

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 2004, 20:58:06 ; Search time 3642 Seconds
(without alignments)
10841.709 Million cell updates/sec

Title: US-10-069-434-4
Perfect score: 911
Sequence: 1 cgggaattcggtcgagtcc.....cccaaatgccagggggaaaa 911

Scoring table: IDENTITY NUC

Gapop 10_0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pi.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pi.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	911	100.0	911	6	AX350366	AX350366 Sequence
2	759	83.3	1248	6	AX552095	AX552095 Sequence
3	745.8	81.9	1023	6	AX458348	AX458348 Sequence
4	745.8	81.9	1237	9	BC052602	BC052602 Homo sapi
5	745.8	81.9	3564	6	AX833515	AX833515 Sequence
6	745.8	81.9	3564	9	AK095314	AK095314 Homo sapi
7	740.6	81.3	828	6	AX817267	AX817267 Sequence
8	728.8	80.0	833	6	AX817269	AX817269 Sequence
9	590	64.8	1052	10	AF231123	AF231123 Mus muscu
10	590	64.8	1082	10	BC064050	BC064050 Mus muscu
c 11	575.2	63.1	267093	2	AC109408	AC109408 Rattus no
12	506.4	55.6	266620	2	AC112526	AC112526 Rattus no
c 13	506.4	55.6	276378	2	AC110854	AC110854 Rattus no
14	493.2	54.1	688	6	BD276051	BD276051 48 Human
15	481.4	52.8	2403	9	AK093753	AK093753 Homo sapi
16	341	37.4	1231	4	MD0417908	AJ417908 Monodelph
17	331.2	36.4	914	9	HSCAIR	X05014 Human CDNA
18	331.2	36.4	2785	9	HUMCAIX	M33987 Human carbo
19	331.2	36.4	2785	11	G28544	G28544 Human carbo
20	329.6	36.2	1307	9	BC027890	BC027890 Homo sapi
21	328	36.0	810	9	GORCAIPRO	L11622 Gorilla gor
22	326.4	35.8	810	9	CHPCAIPRO	L11621 Pan troglod
23	325.2	35.7	1125	9	MACCAANI	L25082 Macaca neme
24	321.6	35.3	1440	4	SHPCAI	L42178 Ovis aries
25	311.8	34.2	1224	10	MUSCAIA	M32452 Mouse carbo
26	310.2	34.1	1195	10	BC011223	BC011223 Mus muscu
27	304.2	33.4	786	5	AY125007	AY125007 Lepisoste
28	297.8	32.7	1976	5	AB055617	AB055617 Tribolodo
29	296.2	32.5	1445	5	GDCARANH	Z14957 G.domesticu
30	294.8	32.4	1486	5	GGCAIIR	X04810 Chicken mRN
31	294.2	32.3	783	5	AY307082	AY307082 Oncorhync
32	291.4	32.0	1199	5	GCAIIR	X12839 Chicken mRN
33	289.4	31.8	1459	6	AX401999	AX401999 Sequence
34	289.4	31.8	1459	10	RNCAII	X58294 R.norvegicu
35	284.8	31.3	605	6	I95751	I95751 Sequence 17
36	284.8	31.3	605	6	AR182211	AR182211 Sequence
37	283.6	31.1	881	4	RABCAI	M10412 Rabbit eryt
38	283.4	31.1	1490	10	MUSCAIIM	K00811 Mouse carbo
39	283.4	31.1	1542	10	BC055291	BC055291 Mus muscu
40	281.8	30.9	1574	5	DRU55177	U55177 Danio rerio
41	280	30.7	780	6	AR111448	AR111448 Sequence
42	280	30.7	864	6	AR077867	AR077867 Sequence
43	280	30.7	867	6	AR159858	AR159858 Sequence
44	280	30.7	867	6	AR216143	AR216143 Sequence
45	280	30.7	1363	9	HUMCAIZ	M36532 Human carbo

ALIGNMENTS

RESULT 1

AX350366
LOCUS AX350366 911 bp
DEFINITION Sequence 4 from Patent WO0200840.
ACCESSION AX350366
VERSION AX350366.1 GI:18616025

KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM Homo sapiens

REFERENCE

1 Thornton, M., Ramkumar, J., Tribouley, C.M., Yue, H., Nguyen, D.B.,
Yao, M.G., Patterson, C., Gandhi, A.R., Burford, N., Thangavelu, K. and
Baughn, M.R.


```
QY 481 CAGATTGGTGAACCTAAATTTCCCACTGCAAAAGATTACTGACACTTTGGATTCCCAATATAA 540
Db 495 CAGATTGGTGAACCTAAATTTCCCACTGCAAAAGATTACTGACACTTTGGATTCCCAATATAA 554
QY 541 GAAAGGGTAAACAACCTGATTACAAATTTTGACCTATTCTGCTCTTCCACCATCC 600
Db 555 GAAAGGGTAAACAACCTGATTACAAATTTTGACCTATTCTGCTCTTCCACCATCC 614
QY 601 TGGGACTACTGACATATCTCTGTTCTCTTACAGTTTCCACCTTCTTCTTGAGAGTGTCA 660
Db 615 TGGGACTACTGACATATCTCTGTTCTCTTACAGTTTCCACCTTCTTCTTGAGAGTGTCA 674
QY 661 TGGATTGTTTTAAAGCAACCTATAAATCAATGAGTCTCAACAGCTGGCCAAATTTGCGAGT 720
Db 675 TGGATTGTTTTAAAGCAACCTATAAATCAATGAGTCTCAACAGCTGGCCAAATTTGCGAGT 734
QY 721 CTCCTGTGCACAGCGGGGTGAAGCAGCAGCTTTTCTG 759
Db 735 CTCCTGTGCACAGCGGGGTGAAGCAGCAGCTTTTCTG 773

RESULT 3
LOCUS AX458348 1023 bp DNA linear PAT 08-JUL-2002
DEFINITION Sequence 19 from Patent WO0246385.
ACCESSION AX458348
VERSION AX458348.1 GI:21725020
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE
AUTHORS Tang,Y.T., Griffin,J.A., Yue,H., Lee,E.A., Baughn,M.R.,
Duggan,B.M., Walla,N.K., Lee,S., Ramkumar,J., Warren,B.A.,
Gandhi,A.R., Lu,D.A., Lu,Y., Yao,M.G., Ding,L., Tribouley,C.M.,
Sanjanwala,M.M., Arvizu,C. and Hillman,J.L.
Enzymes
Patent: WO 0246385-A 19 13-JUN-2002;
Incyte Genomics, Inc. (US)
FEATURES
source
1..1023
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/note="Incyte ID No: 8159895CB1"
ORIGIN
Query Match 81.9%; Score 745.8; DB 6; Length 1023;
Best Local Similarity 97.2%; Pred. No. 1.4e-189;
Matches 759; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
QY 17 TTCCACCCCGAGGGACCATGTGAGGCTCAGCTGGGGATACCGGAGCAACAGGTCCTA 76
Db 1 TTCCACCCCGAGGGACCATGTGAGGCTCAGCTGGGGATACCGGAGCAACAGGTCCTA 60
QY 77 TTCCTCTGGAAGGAATTTTCCCTATTGCTGATGGTGATCAGCAATCTCCAAATTGAGATTA 136
Db 61 TTCCTCTGGAAGGAATTTTCCCTATTGCTGATGGTGATCAGCAATCTCCAAATTGAGATTA 120
QY 137 AAACCAAGAGTGAATATGACTCTTCCCTCGACCATTTAGTATCAATGATGACCCAA 196
Db 121 AAACCAAGAGTGAATATGACTCTTCCCTCGACCATTTAGTATCAATGATGACCCAA 180
QY 197 GCTCAGCTAAATCATCAGCAACAGGGCCATTCCTTCAATGTGACTTTGATGACACAG 256
Db 181 GCTCAGCTAAATCATCAGCAACAGGGCCATTCCTTCAATGTGACTTTGATGACACAG 240
QY 257 AGAACAATAATCAGTTCTGCGTGGTGGTCTCTCACTGGAAAGCTACAGGTTACGCGAGGTTTC 316
Db 241 AGAACAATAATCAGTTCTGCGTGGTGGTCTCTCACTGGAAAGCTACAGGTTACGCGAGGTTTC 300
```

```
QY 317 ACCTTCACTGGGGTCCCGCTGATGACACGGCTCCGAGCATAGTAGTGAAGTGAGCT 376
Db 301 ACCTTCACTGGGGTCCCGCTGATGACACGGCTCCGAGCATAGTAGTGAAGTGAGCT 360
QY 377 ATGCTGCAGAGCTCCATGTTGTTCACTGGAATTCAGACAAATACCCAGCTTTGTTGAGG 436
Db 361 ATGCTGCAGAGCTCCATGTTGTTCACTGGAATTCAGACAAATACCCAGCTTTGTTGAGG 420
QY 437 CAGCTCATGAACACAGATGGAGTGGCTGTCTTGGGAGTGTTTTTACAGATTGGTGAACCTA 496
Db 421 CAGCTCATGAACACAGATGGAGTGGCTGTCTTGGGAGTGTTTTTACAGATTGGTGAACCTA 480
QY 497 ATTCCCAACTGCARAAAGATTACTGACACTTTGGATTCCATTAAGAAAAGGTTAAACAA 556
Db 481 ATTCCCAACTGCARAAAGATTACTGACACTTTGGATTCCATTAAGAAAAGGTTAAACAA 540
QY 557 CTCGATTACAAAATTTTGACCTATTGTTCTGTTCTCCACCATCTCGGACTACTGGACAT 616
Db 541 CTCGATTACAAAATTTTGACCTATTGTTCTGTTCTCCACCATCTCGGACTACTGGACAT 600
QY 617 ATCCTGGTTCCTTACAGTTCACCTCTTCTTGAGAGTGTACATGGATTGTTTAAAGC 676
Db 601 ATCCTGGTTCCTTACAGTTCACCTCTTCTTGAGAGTGTACATGGATTGTTTAAAGC 660
QY 677 AACCTATAAATCAATCAGCTCTCAACAGCTGGCCAAATTTGCGAGTCTCCTGTGCACAGCG 736
Db 661 AACCTATAAATCAATCAGCTCTCAACAGCTGGCCAAATTTGCGAGTCTCCTGTGCACAGCG 720
QY 737 AGGGTGAAGCAGCAGCTTTTCTGTGATAGAGTCTCACTCTGTCAACCCAGGCTGGAGGCA 796
Db 721 AGGGTGAAGCAGCAGCTTTTCTGTGATAGAGTCTCACTCTGTCAACCCAGGCTGGAGGCA 780
QY 797 G 797
Db 781 G 781

RESULT 4
LOCUS BC052602
DEFINITION Homo sapiens similar to carbonic anhydrase 13, mRNA (cDNA clone
ACCESSION BC052602.1 GI:30851208
VERSION BC052602
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE
AUTHORS Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1237)
Strausberg,R.B., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,A.M., Wang,J., Hsieh,F.,
Diatchenko,L., Marusina,K., Farmer,A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Scheet,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettman,M., Madan,A., Rodrigues,S.,
Sanchez,A., Whitting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalau,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
JOURNAL 22388257
MEDLINE 12477932
PUBMED
```


source	1. .3564 /organism="Homo sapiens" /mol_type="unassigned DNA" /db_xref="taxon:9606"	Query Match Best Local Similarity 81.9%; Score 745.8; DB 6; Length 3564; Matches 759; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
ORIGIN	17 TTCCACCCCGAGGACCATGTCGAGGCTCAGCTGGGGATACCGGAGACACACGTCCTA 76 Db 35 TTCCACCCCGAGGACCATGTCGAGGCTCAGCTGGGGATACCGGAGACACACGTCCTA 94 Qy 77 TTCCACCCCGAGGACCATGTCGAGGCTCAGCTGGGGATACCGGAGACACACGTCCTA 136 Db 95 TTCCACCCCGAGGACCATGTCGAGGCTCAGCTGGGGATACCGGAGACACACGTCCTA 154 Qy 137 AAACCAAGAGTGAATATGATCTTCCCTCGGACCATCTAGTATCAAGTATGACCCAA 196 Db 155 AAACCAAGAGTGAATATGATCTTCCCTCGGACCATCTAGTATCAAGTATGACCCAA 214 Qy 197 GCTCAGCTAAATCATCAGCAACACGCGGCATTCCTTCAATGTTGACTTTGATGACACAG 256 Db 215 GCTCAGCTAAATCATCAGCAACACGCGGCATTCCTTCAATGTTGACTTTGATGACACAG 274 Qy 257 AGAACCAATCAGTCTCGGTGGTGCCTCTCACTGGAGCTACAGGTTACGGCAGGTTTC 316 Db 275 AGAACCAATCAGTCTCGGTGGTGCCTCTCACTGGAGCTACAGGTTACGGCAGGTTTC 334 Qy 317 ACCTTCACTGGGGTCCGCTGATGATGACCGGCTCCGAGCACATAGTAGAGTGAGCT 376 Db 335 ACCTTCACTGGGGTCCGCTGATGATGACCGGCTCCGAGCACATAGTAGAGTGAGCT 394 Qy 377 ATGCTGCAGAGCTCCATGTTGTTCACTGGAATTCAGACAAATACCCGACTTTGTGAGG 436 Db 395 ATGCTGCAGAGCTCCATGTTGTTCACTGGAATTCAGACAAATACCCGACTTTGTGAGG 454 Qy 437 CAGCTCATGAACACAGATGGACTGGCTGCTCTGGGAGTGTGTTTACAGATTGGTGAACCTA 496 Db 455 CAGCTCATGAACACAGATGGACTGGCTGCTCTGGGAGTGTGTTTACAGATTGGTGAACCTA 514 Qy 497 ATTCCCAACTGAAAGATATCTGACACTTTGGATTCCATTAAGAAAAGGTAACAAA 556 Db 515 ATTCCCAACTGAAAGATATCTGACACTTTGGATTCCATTAAGAAAAGGTAACAAA 574 Qy 557 CTCGATTCAAAATTTTGACCTATTGCTCTGCTTCCACATCCTGGGACTACTGGACAT 616 Db 575 CTCGATTCAAAATTTTGACCTATTGCTCTGCTTCCACATCCTGGGACTACTGGACAT 634 Qy 617 ATCCTGGTTCCTTACAGTTCACCTCTTCTTGAGAGTGTCAATGGATTGTTTAAAGC 676 Db 635 ATCCTGGTTCCTTACAGTTCACCTCTTCTTGAGAGTGTCAATGGATTGTTTAAAGC 694 Qy 677 AACCTATAAATCAGCTCTCAACAGCTGGCCAAATTTGCGAGTCTCCTGTGTCACAGCG 736 Db 695 AACCTATAAATCAGCTCTCAACAGCTGGCCAAATTTGCGAGTCTCCTGTGTCACAGCG 754 Qy 737 AGGGTGAACGACGAGCTTTTCTGTGATAGTCTCACTCTGTCACCGAGCTGGAGGCA 796 Db 755 AGGGTGAACGACGAGCTTTTCTGTGATAGTCTCACTCTGTCACCGAGCTGGAGGCA 814 Qy 797 G 797 Db 815 G 815	
RESULT 6	AK095314 3564 bp mRNA linear PRI 15-JUL-2002	
LOCUS	Homo sapiens cDNA FLJ37995 fis, clone CTONG2011825, moderately	
DEFINITION	similar to CARBONIC ANHYDRASE (EC 4.2.1.1).	
ACCESSION	AK095314	
VERSION	AK095314.1	GI:21754548
KEYWORDS	oligo capping; fis (full insert sequence).	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE	1 Tanigami, A., Fujiwara, T., Shibahara, T., Goto, Y., Hirao, M., Shimizu, F., Wakebe, H., Ono, T., Hishigaki, H., Watanabe, T., Ozaki, K., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahara, K., Masuho, Y., Nagai, K. and Isogai, T. NEDO human cDNA sequencing project	
JOURNAL	Unpublished	
REFERENCE	2 (bases 1 to 3564)	
AUTHORS	Isogai, T. and Yamamoto, J.	
TITLE	Direct Submission	
JOURNAL	Submitted (04-JUN-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)	
COMMENT	NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.	
FEATURES	Location/Qualifiers 1. .3564 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="CTONG2011825" /tissue_type="tongue, tumor tissue" /clone_lib="CTONG2" /note="cloning vector: pME18SFL3" 52. .840 /note="unnamed protein product" /codon_start=1 /protein_id="BAC04528.1" /db_xref="GI:21754549" /translation="MSRLSWGYREHNGPIHMKFFPIADGDDQSPFIKTKVKYDSS LRPLSIKYDPSAKIISNSGHSFNVDFTENKSVLRGGLTGSYRLRQVHLHWSAD DHGSHIVDGVSYAAELHVVHNSDKPSFVEAAHEPDGLAVLQIGENSOLOK ITDITLDSIKEGKQTRFTNFDLSLLPPSDWYTPGSLTPVPLLESVTWILKQFIN ISSQOLAKFRSLCTAEGEAAAFVLSNHRPPQPLKGRKVRASFH"	
CDS	Query Match 81.9%; Score 745.8; DB 9; Length 3564; Best Local Similarity 97.2%; Pred. No. 1.4e-189; Matches 759; Conservative 0; Mismatches 22; Indels 0; Gaps 0;	
ORIGIN	17 TTCCACCCCGAGGACCATGTCGAGGCTCAGCTGGGGATACCGGAGACACACGTCCTA 76 Db 35 TTCCACCCCGAGGACCATGTCGAGGCTCAGCTGGGGATACCGGAGACACACGTCCTA 94 Qy 77 TTCCACCCCGAGGACCATGTCGAGGCTCAGCTGGGGATACCGGAGACACACGTCCTA 136 Db 95 TTCCACCCCGAGGACCATGTCGAGGCTCAGCTGGGGATACCGGAGACACACGTCCTA 154 Qy 137 AAACCAAGAGTGAATATGATCTTCCCTCGGACCATCTAGTATCAAGTATGACCCAA 196 Db 155 AAACCAAGAGTGAATATGATCTTCCCTCGGACCATCTAGTATCAAGTATGACCCAA 214 Qy 197 GCTCAGCTAAATCATCAGCAACACGCGGCATTCCTTCAATGTTGACTTTGATGACACAG 256 Db 215 GCTCAGCTAAATCATCAGCAACACGCGGCATTCCTTCAATGTTGACTTTGATGACACAG 274 Qy 257 AGAACCAATCAGTCTCGGTGGTGCCTCTCACTGGAGCTACAGGTTACGGCAGGTTTC 316	

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 Decristoforo,M.F., Padigar,M., Miller,C., Tchernev,V., Zhong,H., Zhong,M., Anderson,D., Ballinger,R., Gerlach,V., Spytek,K.A., Rastelli,L., Kekuda,R., Guo,X., Zerhusen,B., Andrew,D., Mezes,P., Patirajan,M., Burgess,C.E., Eisen,A., Wolenc,A., Baungartner,J., Shimkets,R.A., Gusev,V., Vernet,C.A., Taupier,R.J., Pena,C., Shenoy,S., Li,L., Casman,S., Bolgog,F., Fernandes,E., Smithson,G., Malyankar,U., Tallon,B. and Liu,X.

TITLE Novel polypeptides and nucleic acids encoded thereby

JOURNAL patent: WO 02081517-A 17 17-OCT-2002; Curagen Corporation (US)

FEATURES source Location/Qualifiers 1..833 /organism="Homo sapiens" /mol_type="unassigned DNA" /db_xref="taxon:9606" 21..809 /note="unnamed protein product" /codon_start=1 /protein_id="CAE84340.1" /db_xref="GI:39722657" /translation="MSRLSWGRENHPIHKEFPPIADGQDSPIETKVKYDSS LRPSIKYDPSAKIISNSGHSFVDFDPTENKSVLRGPTLGSYRLQVHLHWSAD DHGSHIVDGVSYAELHVVHNSDKYFVEAHPDGLAVGLVFLQIGPNSQLQK ITDLDLSKEKQKTRFNFDPLCLLPSSNDYTYPGSLTPVPLLESVTWILVKQPIN ISSQOLAKFRSLLTAEGEAAFLVSNHRPPQPLKGRKVRASFH"

ORIGIN Query Match 80.0%; Score 728.8; DB 6; Length 833; Best Local Similarity 97.1%; Pred. No. 5.3e-185; Matches 742; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 34 ATGTGAGGCTCAGCTGGGATACCGGAGCACAAGCGTCTTATTCACCTGGAAGAAATTT 93

Db 21 ATGTGAGGCTCAGCTGGGATACCGGAGCACAAGCGTCTTATTCACCTGGAAGAAATTT 80

QY 94 TTCCCTATTGCTGATGCTGATCAGCAATCTCCATTTGATGACACAGACAAATTCAGTTCTG 273

Db 81 TTCCCTATTGCTGATGCTGATCAGCAATCTCCATTTGATGACACAGACAAATTCAGTTCTG 260

QY 154 TATGACTCTTCCTCGGACCTTAGTATCAAGTATGACCCAGCTCAGTAAATCATC 213

Db 141 TATGACTCTTCCTCGGACCTTAGTATCAAGTATGACCCAGCTCAGTAAATCATC 200

QY 214 AGCAACAGCGGCATTCCTTCAATGTTGACTTTGATGACACAGACAAATTCAGTTCTG 273

Db 201 AGCAACAGCGGCATTCCTTCAATGTTGACTTTGATGACACAGACAAATTCAGTTCTG 260

QY 274 CGTGGTGGTCTCTCTACTGGAAGCTACAGTTACGGCAGGTTCCCTTCACTGGGGGTCC 333

Db 261 CGTGGTGGTCTCTCTACTGGAAGCTACAGTTACGGCAGGTTCCCTTCACTGGGGGTCC 320

QY 334 GCTGATGACACGGCTCCGAGCAGATAGTAGTGGAGTGAGTATGCTGCAGAGCTCCAT 393

Db 321 GCTGATGACACGGCTCCGAGCAGATAGTAGTGGAGTGAGTATGCTGCAGAGCTCCAT 380

QY 394 GTTGTTCCTGGAATTCAGACAAATACCCAGCTTTTGTGAGCAGCTCATGACACAGAT 453

Db 381 GTTGTTCCTGGAATTCAGACAAATACCCAGCTTTTGTGAGCAGCTCATGACACAGAT 440

QY 454 GGACTGGTGTCTTGGAGTGTGTTTTACAGATTGGTGAACCTTAATTCCTGTTCTCTTACA 633

Db 441 GGACTGGTGTCTTGGAGTGTGTTTTACAGATTGGTGAACCTTAATTCCTGTTCTCTTACA 620

QY 514 ATTACTGACACTTGGATTCCATTAAAGAAAGGGTAAACAAACTCGATTCAAAATTTT 573

Db 501 ATTACTGACACTTGGATTCCATTAAAGAAAGGGTAAACAAACTCGATTCAAAATTTT 560

QY 574 GACCTATTGCTCTGCTTCCACCATCTCTGGACTACTGGACATATCTGTTCTCTTACA 633

Db 561 GACCTATTGCTCTGCTTCCACCATCTCTGGACTACTGGACATATCTGTTCTCTTACA 620

QY 634 GTTCCACCTCTTCTGAGAGTGTACATGGAGTGTGTTTAAAGCAACCTTAAACATCAGC 693

Db 621 GTTCCACCTCTTCTGAGAGTGTACATGGAGTGTGTTTAAAGCAACCTTAAACATCAGC 680

QY 694 TCTCAACAGCTGGGCAAAATTTCCGAGTCTCTGTGTCACAGCGGAGGTGAAGCAGCAGCT 753

Db 681 TCTCAACAGCTGGGCAAAATTTCCGAGTCTCTGTGTCACAGCGGAGGTGAAGCAGCAGCT 740

QY 754 TTTCTGTGATGAGTCTCCTCTGTGTCACCCAGGCTGGAGGCGAG 797

Db 741 TTTCTGTGATGAGTCTCCTCTGTGTCACCCAGGCTCTTAAAGGCGG 784

RESULT 9 AF231123 1052 bp mRNA linear ROD 20-APR-2001

LOCUS Mus musculus carbonic anhydrase XIII (Carl3) mRNA, complete cds.

DEFINITION AF231123

ACCESSION AF231123

VERSION AF231123.1 GI:13249092

KEYWORDS Mus musculus (house mouse)

SOURCE Mus musculus

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 1052) Hewett-Emmett,D. Evolution and distribution of the carbonic anhydrase gene families (in) Chegwidien,W.R., Carter,N.D. and Edwards,Y.H. (Eds.), THE CARBONIC ANHYDRASES: NEW HORIZONS: 29-76; Birkhauser Verlag, Basel (2000)

AUTHORS Birkhauser Verlag, Basel (2000)

TITLE Evolution and distribution of the carbonic anhydrase gene families

JOURNAL Hewett-Emmett,D. and Shimmin,L.C. Characterization and evolution of two new members of the alpha-carbonic anhydrase gene family in mouse: Carl3 and Carl5 Unpublished

REFERENCE 3 (bases 1 to 1052) Shimmin,L.C. and Hewett-Emmett,D. Direct Submission

AUTHORS Submitted (03-FEB-2000) Human Genetics Center, The University of Texas-Houston School of Public Health, P.O. Box 20334, Houston, TX 77225-0334, USA

TITLE Characterization and evolution of two new members of the alpha-carbonic anhydrase gene family in mouse: Carl3 and Carl5 Unpublished

JOURNAL Shimmin,L.C. and Hewett-Emmett,D. Direct Submission

AUTHORS Submitted (03-FEB-2000) Human Genetics Center, The University of Texas-Houston School of Public Health, P.O. Box 20334, Houston, TX 77225-0334, USA

FEATURES Location/Qualifiers 1..1052 /organism="Mus musculus" /mol_type="mRNA" /strain="C3H" /db_xref="taxon:10090" /clone="IMAGE:1193823" /cell_line="C2C12" /clone_lib="Barstead mouse myotubes MPL-RB5" 1..1052 /gene="Carl3" 33..821 /gene="Carl3" /EC_number="4.2.1.1" /note="orthologous to 8q21.3 Homo sapiens genomic sequence in working draft NT_008117.2 and adjacent to CAL1, CA3 and CA2 genes; this region is syntenic with mouse chromosome 3 region containing Carl1, Carl2 and Carl3" /codon_start=1 /product="carbonic anhydrase XIII" /protein_id="AAK16672.1" /db_xref="GI:13249093" /translation="MARISWGYGEHNGPIHWNELFPIADGQDSPIETKVKYDSS LRPLSKYDPSAKIISNSGHSFVDFDPTENKSVLRGPTLGSYRLQVHLHWSAD DHGSHIVDGVSYAELHVVHNSDKYFVEAHPDGLAVGLVFLQIGPNSQLQK ITDLDLSKEKQKTRFNFDPLCLLPSSNDYTYPGSLTPVPLLESVTWILVKQPIN ISSQOLAKFRSLLTAEGEAAFLVSNHRPPQPLKGRKVRASFH"

gene 1..1052

CDS 33..821

ORIGIN Query Match 64.8%; Score 590; DB 10; Length 1052; Best Local Similarity 84.2%; Pred. No. 1.1e-147; Matches 665; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

AC112526	266620 bp	DNA	linear	HTG 13-MAY-2003	JOURNAL
LOCUS					
DEFINITION	Rattus norvegicus clone CH230-223F17, *** SEQUENCING IN PROGRESS				
AC112526	***, 6 unordered pieces.				COMMENT
AC112526.5	GI:30578729				
VERSION	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.				
KEYWORDS	Rattus norvegicus (Norway rat)				
SOURCE	Rattus norvegicus				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
REFERENCE	1 (bases 1 to 266620)				
AUTHORS	Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogue, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhewa, L., Loulseghe, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokelimeh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkoch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., Puaro, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steidle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villabana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstein, G., and Gibbs, R. A.				
Direct Submission					
Unpublished					
REFERENCE	2 (bases 1 to 266620)				
AUTHORS	Worley, K.C.				
TITLE	Direct Submission				
JOURNAL	Submitted (22-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA				
REFERENCE	3 (bases 1 to 266620)				
AUTHORS	Rat Genome Sequencing Consortium.				
TITLE	Direct Submission				

Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA	
On May 13, 2003 this sequence version replaced gi:25007874.	
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.	
----- Genome Center	
Center: Baylor College of Medicine	
Center code: BCM	
Web site: http://www.hgsc.bcm.tmc.edu/	
Contact: hgsc-help@bcm.tmc.edu	
----- Project Information	
Center project name: GJXR	
Center Clone name: CH230-223F17	
----- Summary Statistics	
Assembly program: Atlas 3.0;	
Consensus quality: 217930 bases at least Q40	
Consensus quality: 22815 bases at least Q30	
Consensus quality: 225895 bases at least Q20	
Estimated insert size: 231511; sum-of-contigs estimation	
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation	

* NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).	
* NOTE: This is a 'working draft' sequence. It currently consists of 6 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.	
* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.	
* 1 94261: contig of 94261 bp in length	
* 94262 94361: gap of unknown length	
* 94362 259486: contig of 165125 bp in length	
* 259487 259586: gap of unknown length	
* 259587 260962: contig of 1376 bp in length	
* 260963 261062: gap of unknown length	
* 261063 262210: contig of 1148 bp in length	
* 262211 262310: gap of unknown length	
* 262311 263690: contig of 1380 bp in length	
* 263691 263790: gap of unknown length	
* 263791 266620: contig of 2830 bp in length.	
FEATURES	
source	Location/Qualifiers
	1..266620
	/organism="Rattus norvegicus"
	/mol_type="genomic DNA"
	/db_xref="taxon:10116"
	/clone="CH230-223F17"
	/complement(2280..3096)
	/note="Clone boundary clone end:Sp6 site:EcoRI
	end sequence:BZ097870"
	93464..94262
	/note="Clone boundary clone end:T7
	site:EcoRI
	end sequence:BZ097867"
	94362..96182
	/note="wgs end extension clone end:T7"
	247612..249655
misc_feature	
misc_feature	
misc_feature	
misc_feature	

Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On May 13, 2003 this sequence version replaced gi:25007874. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GJXR
Center clone name: CH230-223F17
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 217930 bases at least Q40
Consensus quality: 22815 bases at least Q30
Consensus quality: 225895 bases at least Q20
Estimated insert size: 231511; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 94361: contig of 94261 bp in length
94361: gap of unknown length
94362 259486: contig of 165125 bp in length
259487 259586: gap of unknown length
259587 260362: contig of 1376 bp in length
260963 261062: gap of unknown length
261063 26210: contig of 1148 bp in length
26211 262310: gap of unknown length
262311 263690: contig of 1380 bp in length
263691 263790: gap of unknown length
263791 266620: contig of 2830 bp in length.

FEATURES
Location/Qualifiers
1..266620
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-223F17"
complement(2280..3096)
/note="clone boundary
clone_end:Sp6
site:EcORI
end sequence:BZ097870"
93464..94262
/note="clone boundary
clone_end:T7"
misc_feature
misc_feature
misc_feature
misc_feature

ORIGIN		/note="wgs_end_extension clone_end:T7"		SOURCE ORGANISM	
Query Match		55.6%;	Score 506.4;	Rattus norvegicus (Norway rat)	
Best Local Similarity		80.1%;	Pred. No. 3.7e-125;	Rattus norvegicus	
Matches 633;		Conservative 0;	Mismatches 151;	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
			Indels 6;	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;	
			Gaps 3;	Rattus.	
				1 (bases 1 to 276378)	
Qy		6	TTGCGCTCGAGTTCCACCGGACCAATGTCGAGGCTCAGCTGGGGATACCGCGAGCA	Muzny,D.,Marie., Metzker,M.,Lee., Abramson,S., Adams,C., Alder,J.,	
Db		1772	TCTGCGCACAGGTTCTCCATCGAGGACCAATGCGAGCTAAAGCTGGGGATACGGTGAGCA	Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,	
Qy		66	CAACGGTCTATTCACTGCAAGAAATTTTCCCTATTGCTGATGGTGATCAGCAACTCC	Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,	
Db		1832	CAACGGTCTATTCACTGCAAGAAATTTTCCCTATTGCTGATGGTGATCAGCAACTCC	Baldwin,D., Bandaranaike,B., Barber,M., Barnstead,M., Benahmed,F.,	
Qy		126	AATTGAGATTAACCAAGAAAGTGAATATGACTCTTCCCTCCGACCACTTAGTATCAAA	Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,	
Db		1891	AGTTGAGATTAACCAAGAAAGTGAATATGACTCTTCCCTCCGACCACTTAGTATCAAA	Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,	
Qy		186	GTATGACCAAGCTCAGCTAAATCATCAGCAACAGGGCCATCTTCAATGTTGACTT	Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,	
Db		1951	GTATGACCAAGCTCAGCTAAATCATCAGCAACAGGGCCATCTTCAATGTTGACTT	Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,	
Qy		246	TGATGACACAGAGCAACATCAGTTCTGCGTGGTGGTCTCTCACTGGAAGCTACAGGTT	Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,	
Db		2011	TGACGACACGGAGCAACATCAGTTCTCTCTGGAGGTCTCTCATTT-----GCTATAGGTT	Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,	
Qy		306	AGGCGAGGTTCACTTCACTGGGGTCCGCTGATGACCAAGGCTCCGAGCACATAGTAGA	Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,	
Db		2067	CGGCGAGTTCATCTGCACTGGGGTCAGCAGACACCTTGGCTCTGAGTCTGGTAGA	Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,	
Qy		366	TGAGTGAGCTATGCTGCGAGAGCTCCATGTTGTTCTACTGGAATTCAGCAATACCCCGAG	Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,	
Db		2127	TAGAGTGAGGTATGCTGTTGAGTTCATGTTGTTCCACTAGATTTTCAGCAATACCCCGAG	Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,	
Qy		426	CTTTGTTGAGGAGCTCATGAACACAGATGGACTGGCTGCTCTGGGAGTGTTTTCACAGAT	Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,	
Db		2187	TTTGTGAGGAGCTCATGAGTCCGATAGGCTGGCTGCTCTGGGAGTGTTTTCACAGAT	Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,	
Qy		486	TGTTGTAACCTAATTCCTCAACTGCACAAAGAAFTACTGACACTTTGGATTTCATTAAGAAAA	Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,	
Db		2247	TGCGGAACCAATCTCTCACTGCG- AAAGATCATCTGCCATTTTGGATTTCATTAAGAAAA	Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,	
Qy		546	GGGTAAACAACTCGATTCACAAATTTGACCTATTGTTCTGCTCCACCATCTCGGGA	Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hoques,M.,	
Db		2306	GGGTAAACAACTCGATTCACAAATTTGACCTATTGTTCTGCTCCACCATCTCGGGA	Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,	
Qy		606	CTACTGGACATATCTGTTCTCTTACAGTTCACCTCTTCTTGAGAGTGTACATGGAT	Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,	
Db		2366	CTACTGGACATATCTGTTCTCTTACAGTTCACCTCTTCTTGAGAGTGTACATGGAT	Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,	
Qy		666	TGTTTTAAAGCAACCTATTAACATCAGCTCTCAACAGCTGGCCAAATTTCCGAGTCTCT	Kowals,C., Kratt,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,	
Db		2426	TGTTTTAAAGCAACCTATTAACATCAGCTCTTAAAGTTCAGTTCAGTTGTCAGATTCGAAACCTCTC	Liu,J., Liu,W., Liu,X., London,P., Longacre,S., Lopez,J.,	
Qy		726	GTGCACAGCGGAGGTGAAGCAGCAGCTTTTCTGTGATAGAGTCTCACTCTGTCAACCCAG	Lorensuhewa,L., Loulsegged,H., Lozado,R.J., Lu,X., Ma,J.,	
Db		2486	ATGTACAGCTGAGGGTGAGGGTGCAGCTTTTCTGTTTAAGCAATCAACGGTCCGCCACAGCC	Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,	
Qy		786	GCTGGAGGCC 795	Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,	
Db		2546	CTTGAAGGCC 2555	Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,	
RESULT 13					
AC110854/c					
LOCUS					
DEFINITION					
AC110854					
Rattus norvegicus clone CH230-108D22, WORKING DRAFT SEQUENCE, 5					
unordered pieces.					
ACCESSION					
AC110854					
VERSION					
AC110854.5					
GI:24635541					
KEYWORDS					
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE		</			

PF 06-APR-2000 JP 2000611671
PR 09-APR-1999 US 60/128696,14-JAN-2000 US 60/176069 PI
craig a rosen,steven m ruben,george komatsoulis cc
FH key Location/Qualifiers

FEATURES
source
1. 688
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 54.1%; Score 493.2; DB 6; Length 688;
Best Local Similarity 95.7%; Pred. No. 1.2e-121;
Matches 507; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 268 GTTCTGCGTGGTCTCTCACTGGAAGTACAGGTTACGCGAGTTCACTTCACTCG 327
Db 43 GTTCTGCGTGGTCTCTCACTGGAAGTACAGGTTACGCGAGTTCACTTCACTCG 102
QY 328 GGGTCGGCTGATGACACCGCTCCGAGCATAGTAGAGTGGAGTATGCTGCGAGAG 387
Db 103 GGGTCGGCTGATGACACCGCTCCGAGCATAGTAGAGTGGAGTATGCTGCGAGAG 162
QY 388 CTCCATGTTGTTCACTGGAATTCAGACAAATACCCAGGCTTTGTTGAGGCGAGCTCATGAA 447
Db 163 CTCCATGTTGTTCACTGGAATTCAGACAAATACCCAGGCTTTGTTGAGGCGAGCTCATGAA 222
QY 448 CCAGATGGAAGTCTGCTGTTGGAGTGTGTTTACAGATGGTGAACCTAATTCCTCAACATG 507
Db 223 CCAGATGGAAGTCTGCTGTTGGAGTGTGTTTACAGATGGTGAACCTAATTCCTCAACATG 282
QY 508 CAAAGATTACTGACACTTTGGATTCATTAAGAAAGGGTAAACAAACTCGATTTCACA 567
Db 283 CAAAGATTACTGACACTTTGGATTCATTAAGAAAGGGTAAACAAACTCGATTTCACA 342
QY 568 AATTGTTGACTATGTTCTGCTGCTCCACCATCTCGGAGTACTGGAACATATCTGTTCT 627
Db 343 AATTGTTGACTATGTTCTGCTGCTCCACCATCTCGGAGTACTGGAACATATCTGTTCT 402
QY 628 CTTACAGTTCCACCTCTCTGAGAGTGTACATGAGTGTGTTTAAAGCAACTATAAAC 687
Db 403 CTTACAGTTCCACCTCTCTGAGAGTGTACATGAGTGTGTTTAAAGCAACTATAAAC 462
QY 688 ATCAGCTCTCAACAGCTGGCAATTTTCGAGTCTCTCTGTCACAGCGAGGTTGAAGCA 747
Db 463 ATCAGCTCTCAACAGCTGGCAATTTTCGAGTCTCTCTGTCACAGCGAGGTTGAAGCA 522
QY 748 GCAGCTTTTCTGTGATAGTCTCACTCTGTCAACCGAGCTGGAGGCGAG 797
Db 523 GCAGCTTTTCTGTGAGCAATCACCGCCACACAGCTCTTAAAGGCGCG 572

RESULT 15
AK093753
LOCUS 2403 bp mRNA linear PRI 06-SEP-2003
DEFINITION Homo sapiens cDNA FLJ36434 fis, clone THYMU2012002.
ACCESSION AK093753
VERSION AK093753.1 GI:21752680
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS
1
Ishibashi,T., Kanehori,K., Yosida,M., Watanabe,S., Ishida,S.,
Ono,Y., Hotta,T., Hiraoka,S., Murakawa,K., Takiguchi,S.,
Kusano,J., Watanabe,M., Fujimori,K., Tanai,H., Ishida,M.,
Yamashita,H., Chiba,Y., Sugiyama,T., Irie,R., Otsuki,T., Sato,H.,
Ota,T., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,
Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Matsuo,K.,
Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K., Wagatsuma,M.,
Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y.,
Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
Unpublished
2 (bases 1 to 2403)
Isogai,T. and Yamamoto,J.
Direct Submission
Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)

COMMENT
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.

FEATURES
Location/Qualifiers
source
1. 2403
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="THYMU2012002"
/tissue_type="thymus"
/clone_lib="THYMU2"
/note="cloning vector: pME18SFL3"

ORIGIN
Query Match 52.8%; Score 481.4; DB 9; Length 2403;
Best Local Similarity 81.6%; Pred. No. 1.9e-118;
Matches 634; Conservative 0; Mismatches 1; Indels 142; Gaps 1;

QY 70 GGTCTATTCTACGTGGAAGGAATTTTCCCTATTGCTGATGGTGATCAGCAATCTCAATT 129
Db 626 GGTCTATTCTACGTGGAAGGAATTTTCCCTATTGCTGATGGTGATCAGCAATCTCAATT 685
QY 130 GAGATTAAACCAAGAGAGTGAATATGACTCTTCCCTCCGACCCTTAGTATCAAGTAT 189
Db 686 GAGATTAAACCAAGAGAGTGAATATGACTCTTCCCTCCGACCCTTAGTATCAAGTAT 745
QY 190 GACCAAGCTCAGCTAAATCATCAGCAACAGCGGCATTCCTCAATGTCATTGAT 249
Db 746 GACCAAGCTCAGCTAAATCATCAGCAACAGCGGCATTCCTCAATGTCATTGAT 805
QY 250 GACAGAGAAACAAATCAGTCTGCTGGTGTGTCTCTCACTGGAAGCTACAGGTACCG 309
Db 806 GACAGAGAAACAAATCAGTCTGCTGGTGTGTCTCTCACTGGAAGCTACAGGTACCG 865
QY 310 CAGGTTCACTTCACTGGGGGTCCGCTGATGATGACACGGCTCCGAGCACAATAGTAGGA 369
Db 866 CAGGTTCACTTCACTGGGGGTCCGCTGATGATGACACGGCTCCGAGCACAATAGTAGGA 925
QY 370 GTGAGTATGTCAGAGCTCCATGTTGTTCACTGGAATTCAGACAAATACCCAGCTTT 429
Db 926 GTGAGTATGTCAGAGCTCCATGTTGTTCACTGGAATTCAGACAAATACCCAGCTTT 985
QY 430 GTTGAGGAGCTCATGAACAGATGGAGTGGTGTCTTGGGAGTGTGTTTACAG----- 483
Db 986 GTTGAGGAGCTCATGAACAGATGGAGTGGTGTCTTGGGAGTGTGTTTACAGGTGTGC 1045
QY 484 ----- 483
Db 1046 CCCTTTGTACACCCCATCTCTCCATGTGTGTTAACTCATGAAACAGAGAATGCTTCACG 1105
QY 484 ----- 483
Db 1106 GATGGCAACAATTCCTTGTGTCACCAAGATAGGGGTGTGTGGCGCAGGATCCGGGT 1165
QY 484 -----ATTGGTCAACTTAATTTCCCAACTGCAAAAGATTACTGACACTTT 527
Db 1166 GCTTTAAACCTTCAGATTGGTGAACCTAATTTCCCAACTGCAAAAGATTACTGACACTTT 1225
QY 528 GGATTCCTAATAAGAAAGGGTAAACAACTCGATTCAAAATTTTGACCTATTGCTCT 587

Db	1226	GGATTCCATTAAAGAAAGGGTAACAACTCGATTTCACAAATTTTGACCTATTGTCTCT	1285
Qy	588	GCTTCCACCATCCTGGGACTACTGGACATATCCTGGTTCTCTTTACAGTTCCAACCTTTCT	647
Db	1286	GCTTCCACCATCCTGGGACTACTGGACATATCCTGGTTCTCTTTACAGTTCCAACCTTTCT	1345
Qy	648	TGAGAGTGTACATGGATTGTTTTAAAGCAACCTATAAACATCAGCTCTCAACAGCT	704
Db	1346	TGAGAGTGTACATGGATTGTTTTAAAGCAACCTATAAACATCAGCTCTCAACAGGT	1402

Search completed: September 13, 2004, 23:48:37
Job time : 3648 secs

Result No.	Score	Query Match	Length	DB	ID	Description
1	728.8	80.0	789	29	AY409020	Homo sapi
2	722.4	79.3	789	29	AY409021	Pan trogl
3	615.4	67.6	620	12	BM767035	K-EST0049
4	590	64.8	1661	11	AK010166	Mus muscu

REFERENCE	AUTHORS	TITLE
2 (bases 1 to 789)	Clark, A.G., Glanowski, M.A., Tanenbaum, S., Wang, G. and Adams, M.D.	Direct Submission

JOURNAL	Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA									
COMMENT	This sequence was made by sequencing genomic exons and ordering them based on alignment.									
FEATURES	Location/Qualifiers									
source	1..789									
	/organism="Homo sapiens"									
	/mol_type="genomic DNA"									
gene	<1..>789									
ORIGIN	/locus_tag="HCM3418"									
Query Match	80.0%; Score 728.8; DB 29; Length 789;									
Best Local Similarity	97.1%; Pred. No. 2.1e-200;									
Matches	742; Conservative 0; Mismatches 22; Indels 0; Gaps 0;									
QY	34	ATGTCGAGGCTCAGCTGGGATACCGGAGCAACCGGTCTTATTCACCTGGAAGGAATTT	93							
Db	1	ATGTCGAGGCTCAGCTGGGATACCGGAGCAACCGGTCTTATTCACCTGGAAGGAATTT	60							
QY	94	TTCCCTATTGCTGATGGTATCAGCAATCTCCAATTGAGATTAAACCCAAAGAGTGAAA	153							
Db	61	TTCCCTATTGCTGATGGTATCAGCAATCTCCAATTGAGATTAAACCCAAAGAGTGAAA	120							
QY	154	TATGACTCTTCCCTCGACCACTTAGTATCAAGTATGACCCAAAGCTCAGCTAAATCATC	213							
Db	121	TATGACTCTTCCCTCGACCACTTAGTATCAAGTATGACCCAAAGCTCAGCTAAATCATC	180							
QY	214	AGCAACAGCGGCAATCTTCAATGTTGACTTTCAGTATGACCAAGCTCAGTAAATCATC	273							
Db	181	AGCAACAGCGGCAATCTTCAATGTTGACTTTCAGTATGACCAAGCTCAGTAAATCATC	240							
QY	274	CGTGGTGGTCTCTCACTGGAAGCTACAGGTTACGGCAGGTTCCACCTTCACTGGGGTCC	333							
Db	241	CGTGGTGGTCTCTCACTGGAAGCTACAGGTTACGGCAGGTTCCACCTTCACTGGGGTCC	300							
QY	334	GCTGATGACACGGGTCGAGCACAATAGTAGAGTGAGTATGCTGAGAGCTCCAT	393							
Db	301	GCTGATGACACGGGTCGAGCACAATAGTAGAGTGAGTATGCTGAGAGCTCCAT	360							
QY	394	GTTGTTTCACTGGAGTGTGTTTACAGATTGGTGAACCTTAATCCCAACTGCAAAAG	453							
Db	361	GTTGTTTCACTGGAGTGTGTTTACAGATTGGTGAACCTTAATCCCAACTGCAAAAG	420							
QY	454	GGACTGGCTGTCTTGGGAGTGTGTTTACAGATTGGTGAACCTTAATCCCAACTGCAAAAG	513							
Db	421	GGACTGGCTGTCTTGGGAGTGTGTTTACAGATTGGTGAACCTTAATCCCAACTGCAAAAG	480							
QY	514	ATTACTGACACTTTGGATTCCATTAAGAAAGGGTAAACAACTCGATTCACAAAATTTT	573							
Db	481	ATTACTGACACTTTGGATTCCATTAAGAAAGGGTAAACAACTCGATTCACAAAATTTT	540							
QY	574	GACCTATTGCTCTGCTTCCACCATCTGGGACTACTGGACATATCTGTTCTCTTACA	633							
Db	541	GACCTATTGCTCTGCTTCCACCATCTGGGACTACTGGACATATCTGTTCTCTTACA	600							
QY	634	GTTCCACCTCTTCTGAGAGTGTACATGATTGTTTAAAGCAACCTATAACATCAGC	693							
Db	601	GTTCCACCTCTTCTGAGAGTGTACATGATTGTTTAAAGCAACCTATAACATCAGC	660							
QY	694	TCTCAACAGCTGGCCAAATTTCCAGTCTCTGTCACAGCGGAGGGTGAAGCAGAGCT	753							
Db	661	TCTCAACAGCTGGCCAAATTTCCAGTCTCTGTCACAGCGGAGGGTGAAGCAGAGCT	720							
QY	754	TTTCTGTGATAGAGTCTCACTCTGTCAACCCAGGCTGGAGGGCAG	797							
Db	721	TTTCTGTGATAGAGTCTCACTCTGTCAACCCAGGCTGGAGGGCAG	764							
RESULT 2										
AY409021										
LOCUS	AY409021	789 bp	DNA	linear	GSS	12-DEC-2003				

DEFINITION	Pan troglodytes HCM3418 gene, VIRTUAL TRANSCRIPT, partial sequence, genomic survey sequence.									
ACCESSION	AY409021									
VERSION	AY409021.1									
KEYWORDS	GSS.									
SOURCE	Pan troglodytes (chimpanzee)									
ORGANISM	Pan troglodytes									
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.									
AUTHORS	1 (bases 1 to 789) Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.									
TITLE	Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios									
JOURNAL	Science 302 (5652), 1960-1963 (2003)									
PUBMED	14671302									
REFERENCE	2 (bases 1 to 789) Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.									
AUTHORS	Direct Submission Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA									
TITLE	This sequence was made by sequencing genomic exons and ordering them based on alignment.									
JOURNAL										
COMMENT										
FEATURES	Location/Qualifiers									
source	1..789 /organism="Pan troglodytes" /mol_type="genomic DNA" /db_xref="taxon:9598" <1..>789 /locus_tag="HCM3418"									
gene										
ORIGIN	Query Match 79.3%; Score 722.4; DB 29; Length 789; Best Local Similarity 96.6%; Pred. No. 1.5e-198; Matches 738; Conservative 0; Mismatches 26; Indels 0; Gaps 0;									
QY	34	ATGTCGAGGCTCAGCTGGGGATACCGGAGCAACCGGTCTTATTCACCTGGAAGGAATTT	93							
Db	1	ATGTCGAGGCTCAGCTGGGGATACCGGAGCAACCGGTCTTATTCACCTGGAAGGAATTT	60							
QY	94	TTCCCTATTGCTGATGGTATCAGCAATCTCCAATTGAGATTAAACCCAAAGAGTGAAA	153							
Db	61	TTCCCTATTGCTGATGGTATCAGCAATCTCCAATTGAGATTAAACCCAAAGAGTGAAA	120							
QY	154	TATGACTCTTCCCTCGACCACTTAGTATCAAGTATGACCCAAAGCTCAGTAAATCATC	213							
Db	121	TATGACTCTTCCCTCGACCACTTAGTATCAAGTATGACCCAAAGCTCAGTAAATCATC	180							
QY	214	AGCAACAGCGGCATTCCTTCAATGTTGACTTTGATGACACAGAGAACAAATCAGTTCTG	273							
Db	181	AGCAACAGCGGCATTCCTTCAATGTTGACTTTGATGACACAGAGAACAAATCAGTTCTG	240							
QY	274	CGTGGTGGTCTCTCACTGGAAGCTACAGGTTACGGCAGGTTTCACCTTCACTGGGGTCC	333							
Db	241	CGTGGTGGTCTCTCACTGGAAGCTACAGGTTACGGCAGGTTTCACCTTCACTGGGGTCC	300							
QY	334	GCTGATGACACAGGCTCCGAGCATATAGTAGAGTGAGTGAGCTATGCTGCAGAGCTCCAT	393							
Db	301	GCTGATGACACAGGCTCCGAGCATATAGTAGAGTGAGTGAGCTATGCTGCAGAGCTCCAT	360							
QY	394	GTTGTTTCACTGGAATTCAGACAAATACTCCAGAGCTTTGTTGAGGAGCTCATGAAACCAT	453							
Db	361	GTTGTTTCACTGGAATTCAGACAAATACTCCAGAGCTTTGTTGAGGAGCTCATGAAACCAT	420							
QY	454	GGACTGGCTCTCTGGAGTGTGTTTTACAGATTGGTGAACCTTAATCCCAACTGCAAAAG	513							
Db	421	GGACTGGCTCTCTGGAGTGTGTTTTACAGATTGGTGAACCTTAATCCCAACTGCAAAAG	480							


```

QY 514 ATTACTGACACTTGGATTCCATTAAAGAAAGGGTAAACAACTCGATTCCAAATTTT 573
Db 481 ATTACTGACACTTGGATTCCATTAAAGAAAGGGTAAACAACTCGATTCCAAATTTT 540
QY 574 GACCTATTCTCTCTGCTTCCACCATCTCTGGGACTACTGGACATATCTCTGTTCTCTTACA 633
Db 541 GACCCATTCTCTCTGCTTCCACCATCTCTGGGACTACTGGACATATCTCTGTTCTCTTACA 600
QY 634 GTTCCACCTCTCTCTGAGAGTGTCACATGAGATTGTTTAAAGCAACCTATAACATCAGC 693
Db 601 GTTCCACCTCTCTCTGAGAGTGTCACATGAGATTGTTTAAAGCAACCTATAACATCAGC 660
QY 694 TCTCAACAGCTGSCCAAAATTCGCACTCTCTGTGACAGCGGAGGTCAGCAGCAGCT 753
Db 661 TCTCAACAGCTGSCCAAAATTCGCACTCTCTGTGACAGCGGAGGTCAGCAGCAGCT 720
QY 754 TTTCTGTGATGAGTCTCACTCTGTGACCCAGGCTGGAGGCGAG 797
Db 721 TTTCTGTGATGAGTCTCACTCTGTGACCCAGGCTGGAGGCGAG 764

RESULT 3
BM767035 620 bp mRNA linear EST 04-MAR-2002
LOCUS K-EST0049075 S14K402 Homo sapiens cDNA clone S14K402-2-F11 5', mRNA
DEFINITION sequence.
VERSION BM767035
KEYWORDS BM767035.1 GI:19096650
SOURCE EST.
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1. (bases 1 to 620)
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
21C Frontier Korean EST Project 2001
Unpublished (2002)
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsu@mail.kribb.re.kr
Plate: 2 row: F column: 11
High quality sequence stop: 620.
Location/Qualifiers
1..620
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="S14K402-2-F11"
/cell_line="K402"
/lab_host="Top10F"
/clone_lib="S14K402"
/note="Organ: Stomach; Vector: pTZ19RP1; Site 1: EcoRI;
Site 2: NotI; The poly (A) + RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then decapped
with tabacco acid pyrophosphatase (TAP). The decapped
intact mRNA was ligated with DNA-RNA linker including EcoR
I site by treatment of T4 RNA ligase and the first strand
cDNA was synthesized from oligo dt-selected mRNA by
priming with dt-tailed vector. The dt-tailed vector was
adjusted to have about 60nt. The cDNA vector was
circularized with E. coli DNA ligase after digestion of
EcoRI which site is also included in vector. An RNA strand
converted to a DNA strand by Okayama-Berg method. The
obtained cDNA vectors were used for transformation of
competent cells E. coli Top10F by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library."

FEATURES
source

```

ORIGIN

```

Query Match 67.6%; Score 615.4; DB 12; Length 620;
Best Local Similarity 99.7%; Pred. No. 1.8e-167;
Matches 616; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 15 AGTTCCACCCCGAGGACCATGTGAGGCTGAGCTGGGATACCGGAGACACACGGTCC 74
Db 3 ACTTCCACCCCGAGGACCATGTGAGGCTGAGCTGGGATACCGGAGACACACGGTCC 62
QY 75 TATTCACTGGAAGGAATTTTCCCTATTGCTGATGATGATCAATCTCCTCAATTTGAGAT 134
Db 63 TATTCACTGGAAGGAATTTTCCCTATTGCTGATGATGATCAATCTCCTCAATTTGAGAT 122
QY 135 TAAACCAAGGAAGTGAATATGACTCTTCCCTCCGACCACTTAGTATCAAGTATGACCC 194
Db 123 TAAACCAAGGAAGTGAATATGACTCTTCCCTCCGACCACTTAGTATCAAGTATGACCC 182
QY 195 AAGCTCAGCTAAATCATCAGCAACAGCGGCCATTCCTCAATGTTGACTTTGATGACAC 254
Db 183 AAGCTCAGCTAAATCATCAGCAACAGCGGCCATTCCTCAATGTTGACTTTGATGACAC 242
QY 255 AGAGAACAAATCAGTTCTGCGTGGTGGTCTCTCACTGGAAGCTACAGGTTACGGCAGGT 314
Db 243 AGAGAACAAATCAGTTCTGCGTGGTGGTCTCTCACTGGAAGCTACAGGTTACGGCAGGT 302
QY 315 TCACCTTCACTGGGGTCCGCTGATGACCAACCGCTCCGAGCACATAGTAGAGTGGAG 374
Db 303 TCACCTTCACTGGGGTCCGCTGATGACCAACCGCTCCGAGCACATAGTAGAGTGGAG 362
QY 375 CTATGCTGAGAGCTCCATGTTGTTCACTGGAAATTCAGACAAATACCCAGCTTTGTTGA 434
Db 363 CTATGCTGAGAGCTCCATGTTGTTCACTGGAAATTCAGACAAATACCCAGCTTTGTTGA 422
QY 435 GGCAGTCAATGAACAGATGGAGTGGCTGCTCTTGGGAGTGTCTTACAGATTGGTGAACC 494
Db 423 GGCAGTCAATGAACAGATGGAGTGGCTGCTCTTGGGAGTGTCTTACAGATTGGTGAACC 482
QY 495 TAAATCCCAACTGCAAAAGATTACTGACACTTTGGATTCCATTAAAGAAAGGGTAAACA 554
Db 483 TAAATCCCAACTGCAAAAGATTACTGACACTTTGGATTCCATTAAAGAAAGGGTAAACA 542
QY 555 AACTGATTCACAAATTTGACCTATTGCTCTGCTCCACCATCTGGGACTACTGGAC 614
Db 543 AACTGATTCACAAATTTGACCTATTGCTCTGCTCCACCATCTGGGACTACTGGAC 602
QY 615 ATATCTGTTCTCTCTTAC 632
Db 603 ATATCTGTTCTCTCTTAC 620

```

RESULT 4

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL


```

QY      788 TGAAGGGCAG 797
      |||||
Db      791 TGAAGGGCG 800

RESULT 5
AY409022
LOCUS
DEFINITION Mus musculus HCM3418 gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION AY409022
VERSION AY409022.1 GI:39764990
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 789)
AUTHORS Clark,A.G., Gnanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 789)
AUTHORS Clark,A.G., Gnanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment
FEATURES
    source
        1..789
            /organism="Mus musculus"
            /mol_type="genomic DNA"
            /db_xref="taxon:10090"
            <1..>789
            /locus_tag="HCM3418"
    gene
    ORIGIN
        Query Match      64.2%; Score 584.8; DB 29; Length 789;
        Best Local Similarity 85.3%; Pred. No. 1.6e-158;
        Matches 652; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY      34 ATGTGAGGCTCAGCTGGGGATACCGCGAGCAACGGTCCATTATTCACCTGGAGGAAATTT 93
      |||||
Db      1 ATGGCGAGGCTAGCTGGGGGTACGGCGAGCACACGGTCCCATTCACCTGGAAATGAATTG 60

QY      94 TTCCTATTGCTGATGGTGATCAGCAATCTCCAAATTGAGATTAAACCAAGAGGTGAAA 153
      |||||
Db      61 TTCCTATTGCTGATGGTGATCAGCAATCTCCAAATTGAGATTAAACCAAGAGGTGAAA 120

QY      154 TATGACTCTTCCTCCGACCATTTAGTATCAAGTATGACCAAGCTCAGCTAAATCATC 213
      |||||
Db      121 TACGACTCTCTACCTCGACCTCTCAGTATCAAGTATGATGCTCGCTCAGCTAAATCATC 180

QY      214 AGCAACAGCGGCCATTCCTTCAATGTTGATGTTGATGACACAGAGAACAAATCAGTTCTG 273
      |||||
Db      181 AGCAATAGTGGCCATTCCTTCAACGTTGACTTTGACACAGCGAGGACAAATCAGTTCTG 240

QY      274 CGTGTGGTCTCTCTACTGGAAGCTACAGGTTAGGCGAGGTTCACTTCACTGGGGGTCC 333
      |||||
Db      241 CGTGGAGGTCCTCTCACTGGGAACTACAGGTTGCGGCGAGTTCCATTGCACTGGGGGTCA 300

QY      334 GCTGATGACACCGGCTCCGAGCACATAGTAGTGGAGTGAGCTATGCTGCAGAGCTCCAT 393
      |||||
Db      301 GCAGATGATCATGGCTCAGAGCATGTGGTAGCGGAGTGAGGTATGCTGCAGAGCTGCAT 360

```

```

QY      394 GTTGTTCATCGAATTACAGACAAATACCCAGCTTTTGTGAGGAGCTCATGAAACAGAT 453
      |||||
Db      361 GTTGTCCACTGGAATTACAGACAAATACCCGAGCTTTGTAGAGGCGACTCATGAGTCCGAT 420

QY      454 GGACTGGCTGTCTTGGGAGTGTGTTTACAGATTGGTGAACCTTAATTCCTCAACTGCAAAAG 513
      |||||
Db      421 GGGCTGGCTGTCTTGGGAGTATTCTACAGATTGGGGAACACAATCCTCAACTGCAAAAG 480

QY      514 ATTACTGCACATTTTGGATTTCATTAAGAAAAGGTAAACAAAACCTCGATTCAAAATTTT 573
      |||||
Db      481 ATCACTGATATTGTTGGATTTCATTAAGAAAAGGTAAACAAAACCTCGATTCAAAATTTT 540

QY      574 GACCTATTGTCTCTGCTTCCACCATCTCGGAGCTACTGGACATATCCTGGTCTCTTACA 633
      |||||
Db      541 GACCCGTTATGCTTGCCTTCCATCATCTTGGGAGCTACTGGACATATCCTGGGCTCTCTGACA 600

QY      634 GTTCCACCTCTTCTTGAGAGTGTACATGGATTGTTTAAAGCAACCTATAAATCAGC 693
      |||||
Db      601 GTTCCACCTCTTCTTGAGAGTGTACATGGATTGTTTAAAGCAACCTATAAATCAGC 660

QY      694 TCTCAACAGCTGGCCAAATTTCCGAGTCTCTGTGTCACAGCGGAGGGTGAACAGCAGCT 753
      |||||
Db      661 TCTCAACAGCTGGCCAGATTCCGAAAGCCTCTTGTGCACAGCTGAGGAGAGAGTGGGCT 720

QY      754 TTTCTGTGATAGAGTCTCACCTCTGTCAACCCAGCTGGAGGGCAG 797
      |||||
Db      721 TTTTGTTAAGCAATCACCGTCCACACAGCCCTTTGAAGGGCGG 764

RESULT 6
CB317335
LOCUS
DEFINITION CB317335
ACCESSION CB317335
VERSION CB317335.1 GI:28841570
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 751)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. Leslie L. Heckert
cDNA Library Preparation: CLONTECH Laboratories, Inc.
DNA Sequencing Arrayed by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NDCM43 row: e column: 06
High quality sequence stop: 539.
Location/Qualifiers
    1..751
        /organism="Mus musculus"
        /mol_type="mRNA"
        /db_xref="taxon:10090"
        /clone="IMAGE:30282437"
        /tissue_type="primary cultures of Sertoli cells"
        /lab_host="DH10B (T1-phage-resistant)"
        /clone_lib="NIH MGC 165"
        /note="Organ: testis; Vector: pDNR-LIB; Site_1: Sfil
        (ggccatctggcc); Site_2: Sfil (ggccgctcgcc); 5' and 3'
        adaptors were used in cloning as follows: 5' adaptor
        sequence: 5'-CAGGCGCAATTATGGC-3' and 3' adaptor sequence:
        5'-ATTCTAGAGCGCGGCGGCACATG-dt(30)BN-3' (where B = A,
        C, or G and N = A, C, G, or T). Average insert size 1.4 kb
        (range 0.6-3.5 kb). 15/15 colonies contained inserts by

```

PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."

ORIGIN	Query Match Best Local Similarity Matches 630; Conservative 0; Mismatches 100; Indels 2; Gaps 2;
QY	29 GGACCATGTCGAGGCTCAGCTGGGATACCGGAGCACAAACGGTCTATTCTACCTGAAGG 88
Db	20 GGACCATGCGAGGCTGAGCTGGGGTACGGCGAGCACAAACGGTCTCATTCTCAGTGAATG 79
QY	89 AATTTTCCCTATTGCTGATGCTGATCAGCAATCTCCAAATTCGATTAACCAAGAG 148
Db	80 AATTTGTTCCCTATTGCTGATGCTGATCAGCAATCTCCAAATTCGATTAACCAAGAG 139
QY	149 TGAATATGACTCTTCCCTCCGACCACTTAGTATCAAGTATGACCCAAAGCTCAGCTAAAA 208
Db	140 TGAATATGACTCTTCCCTCCGACCTCTCAGTATCAAGTATGATCTTGCTCAGCTAAAA 199
QY	209 TCATCAGCAACGCGGCGCATCTTCAATGTTGACTTTTGATGACACAGAGCAAAATCAG 368
Db	200 TCATCAGCAATAGTGCGCATCTTCAAGTTCGACTTTGACGACAGGAGCAAAATCAG 259
QY	269 TTCTGGGTGGTGGTCTCTCACTGGAAGCTACAGGTTACGGCAGGTTCACTTCACTGG 328
Db	260 TTCTGGGTGGGTCCTCTCACTGGAAGCTACAGGTTGCGGAGTTCCATTTGCACTGG 319
QY	329 GGTCCGCTGATGACCAACGCGCTCCGAGCACATAGTAGAGTAGAGTCTGTGTCAGAGC 388
Db	320 GGTGACGATGATCATGGCTCAGAGCATGTGTAGACGAGTGAGTATGCTGTCAGAGC 379
QY	389 TCATGTTGTTCACTGGAAATTCAGAAATACCCCAAGCTTTGTTGAGGAGCTCATGAAC 448
Db	380 TGCAATGTTCTCACTGGAAATTCAGAAATACCCCAAGCTTTGTTGAGGAGCTCATGAGT 439
QY	449 CAGATGACCTGGCTGCTTGGAGTCTTTTACAGATTTGTTGAACTTAATTCCTCACTGC 508
Db	440 CGATGGGCTGGCTGCTTGGAGTATTTCTACAGATTTGGGAAACAAATCTCTCACTGC 499
QY	509 AAAAGATTACTGACACTTTGGATTCCATTAAGAAAGGGTAAACAAACTCGATTACAA 568
Db	500 AAAAGATCACTGATATTTGGATTCCATTAAGAAAGGGTAAACAAACAGGTTCAACA 559
QY	569 ATTTGACCTATGCTCTGCTTCCACATCTCGGAGCTACTGGAATATCTCGTGTCTC 628
Db	560 ATTTGACCGGTTATGCTTCTTCCATCTCTTGGGACTACTGACATACCTCGGCTCTC 619
QY	629 TTACAGTTCACCTCTTCTGAGAGTGCACAT-GGATTGTTTTTAAAGCAACCTATAAAC 687
Db	620 TGACGTTCCACCTCTCTTGAAGTGTACGTGGGATTTGTTTTTAAACACCTATAAGT 679
QY	688 ATCAGCTCTCAACAGCTGGCAAAATTT-GCAGTCTCTGTGTGACAGCGAGGCTGAAGC 746
Db	680 ATCAGCTCTCAACAGCTGGCAGATTTCCGAAACCTCTTGTGTCACAGCTGAAGGAGAG 739
QY	747 AGCAGCTTTTCT 758
Db	740 TGGCGGCTTTT 751

RESULT 7
BQ924401
LOCUS BQ924401
DEFINITION AGENCOURT_8824325 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:6459353
5', mRNA sequence.
ACCESSION BQ924401
VERSION BQ924401.1 GI:22339432
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 975)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/Drp/Gazdar
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LCM2644 row: e column: 18
High quality sequence stop: 661.

FEATURES
source

Location/Qualifiers
1..975
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6459353"
/tissue_type="large cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 18"
/note="Organ: lung; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
Laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."

ORIGIN

Query Match Best Local Similarity Matches 635; Conservative 0; Mismatches 43; Indels 12; Gaps 6;	59.0%; Score 537.2; DB 13; Length 975; 92.0%; Pred. No. 1.3e-144; 0; Mismatches 43; Indels 12; Gaps 6;
QY	17 TTCCACCCCGAGGAGCCATGTCGAGCTCAGCTGGGGATACCGGAGCAACCGTCTCTTA 76
Db	286 TTCCACCCCGAGGAGCCATGTCGAGCTCAGCTGGGGATACCGGAGCAACCGTCTCTTA 345
QY	77 TTCACTGGAAGAAATTTTCCCTATTGCTGATGTCAGCAATCTCCAATTCAGATT 136
Db	346 TTCACTGGAAGAAATTTTCCCTATTGCTGATGTCAGCAATCTCCAATTCAGATT 405
QY	137 AAACCAAAAGAGTGAATATGATCTTCCCTCCGACCACTTAGTATCAAGTATGACCCAA 196
Db	406 AAACCAAAAGAGTGAATATGATCTTCCCTCCGACCACTTAGTATCAAGTATGACCCAA 465
QY	197 GCTCAGCTAAATCATCAGCAACAGCGGCATCTTCAATGTTGATGACACAG 256
Db	466 GCTCAGCTAAATCATCAGCAACAGCGGCATCTTCAATGTTGATGACACAG 525
QY	257 AGAACAATACTGCTTCTGCTGGTCTCTCACTGGAAGCTACAGGTTACGGCAGGTT 316
Db	526 AGAACAATACTGCTTCTGCTGGTCTCTCACTGGAAGCTACAGGTTACGGCAGGTT 585
QY	317 ACCTTCACCTGGGGTCCGCTGATGACCAAGCTCCGAGCAATAGTAGAGTGGAGT 376
Db	586 ACCTTCACCTGGGGTCCGCTGATGACCAAGCTCCGAGCAATAGTAGAGTGGAGT 645
QY	377 ATGCTCAGAGCTCCATGTTGTTCTCACTGGAATTCAGCAAAATACCCAGCTTTGTGAGG 436
Db	646 ATGCTCAGAGCTCCATGTTGTTCTCACTGGAATTCAGCAAAATACCCAGCTTTGTGAGG 705
QY	437 CAGCTCATGAACACAGATGGACTGGCTGTCTTGGAGTGTGTTTTTACA-GATTGCTGAACCT 495
Db	706 CAGCTCATGAACACAGATGGACTGGCTGTCTTGGAGTGTGTTTTTACAGGATGCTGAACCT 765
QY	496 AATTCCCAACTGCAAAAGATTACTGA-CACTTTGGATTCCATTAAAGAAAGGTTAAACA 554

```

Db      766 AATTTCCTCCACTGCCAAAGATTACTGACCCCTTGTGATTCCTATTAAGAAAGGGTAACA 825
Qy      555 AACTCGATTCAAAATTTTGACCTATTGTCTGTG-CTTCCACCAATCCTGGG-----ACTAC 609
Db      826 AACTCGAATTTCCAAATTTTGACCTATGGCTGTGCTTCCACCAATCCTGGGACTACTGG 885
Qy      610 TGGACATATCTCGTGTCTCTTACAGTTCACCC--TCTTCTTGAGAGTGTCCATCGGATT- 666
Db      886 AACCTATCCCGGGTTTCTCTTACGGTCCCCCTCTTTTGGAAAGGGTCCCAATTGAATT 945
Qy      667 --GTTTAAAGCAACTATAAATCATCAGCT 694
Db      946 GGTTTAAAGCCATCTATTAACTCAGCT 975

RESULT 8
LOCUS   BY744350
DEFINITION BY744350 RIKEN full-length enriched, bone marrow macrophage Mus
ACCESSION BY744350
VERSION   BY744350.1 GI:27170890
KEYWORDS EST.
SOURCE   Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE
AUTHORS Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I.,
Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Sult, C.,
Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,
Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V.,
Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,
Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,
Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,
Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,
Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,
Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,
Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
Numata, K., Okido, T., Pavan, W.J., Perte, G., Pesole, G.,
Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S.,
Ravasi, T., Reed, J.C., Reid, J., Reid, J., Ring, B.Z., Ringwald, M.,
Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K.,
Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,
Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,
Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,
Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,
Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,
Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,
Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,
Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,
Rogers, J., Birney, E. and Hayashizaki, Y.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
22354683
12466851
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp
URL: http://genome.gsc.riken.go.jp/
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Hirozane, T., Hori, F.,
Fukuda, S., Hashizume, W., Hayashida, K., Hirozane, T., Kawai, J.,
Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y.,
Kondo, S., Konno, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M.,

```

```

Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N.,
Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M.,
Takeda, Y., Waki, K., Watanabe, A., Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Tissues were provided by David A. Hume ( Depts. of Biochemistry
and Microbiology/Parasitology Institute for Molecular Bioscience
University of Queensland Brisbane, Q 4072 Australia ) whose
assistance we gratefully acknowledge.
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
location/Qualifiers
1. 659
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="I830031F03"
/tissue_type="bone marrow"
/cell_type="macrophage"
/clone_lib="RIKEN full-length enriched, bone marrow
macrophage"

ORIGIN
Query Match 56.0%; Score 510.4; DB 13; Length 659;
Best Local Similarity 85.7%; Pred. No. 6.5e-137;
Matches 565; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 14 GAGTTCCACCCGAGGGACCATGTGAGCTCAGCTGGGGATACCGCGAGCACAGGTC 73
DB 1 GAGGCTCTCCATCGAGGACCATGGCGAGCTGAGCTGGGGGTACGGCGAGCACAGGTC 60
QY 74 CTATTCACTGGGAAGGAATTTTCCCTATTGCTGATGGTGATCAGCAATCTCCAATTGAGA 133
DB 61 CCATTCACTGGGAATGAATGTTCCCTATTGCTGATGGTGATCAGCAATCTCCAATTGAGA 120
QY 134 TTTAAACCAAAAGAGTGAATATGACTCTTCCTCCGACCATCTAGTATCAAGTATGACC 193
DB 121 TTTAAACCAAAAGAGTGAATATGACTCTCTCACTCCGACCTCTCAGTATCAAGTATGATC 180
QY 194 CAAGCTCAGTAAATCATCAGCAACAGCGCCATCTCTCAATGTTGACTTTGATGACA 253
DB 181 CTGCGCTCAGTAAATCATCAGCAATAGTGGCCATCTCTCAACGTTGACTTTGAGGACA 240
QY 254 CAGAGAACAAATCAGTTCTGCGTGGTGGTCTCTCACTGGAAGCTACAGGTTACGGCAGG 313
DB 241 CGAGAGCAAAATCAGTTCTGCGTGGAGGTCTCTCACTGGGAACCTACAGGTTGCGGCAGT 300
QY 314 TTTCACTTTCACCTGGGGGTCCCGTGTATGATGACCAACGGCTCCGAGACATAGTAGTGA 373
DB 301 TCCATTTCACCTGGGGGTCCCGTGTATGATGATCATGGCTCAGAGCATGTGGTAGCGGAGTGA 360
QY 374 GCTATGCTCGAGCTCCATGTTGTTTCACTGGAATTCAGACAAAATCCCGAGCTTTGTTG 433
DB 361 GGTATGCTCGAGAGCTGCAATGTTGTCCACTGGAATTCAGACAAAATCCCGAGCTTTGTTG 420
QY 434 AGGAGCTCATGACCAACAGATGAGCTGGCTGTCTTGGAGTGTGTTTACAGATTGGTGAAC 493

```

Db 421 AGCGAGCTCATGAGTCCGATGGCTGGCTGCTCTGNGAGTATTTCTACAGATTGGGGAAC 480
 QY 494 CTAATTTCCAACTGCAAAAGATTACTGACACTTTGGATTCCATTAAGAAAAGGGTAAAC 553
 Db 481 ACAATCTCTCACTGCAAAAGATCACTGATATTNTGGGATTCTATAAGAGANAAGGGTAAAC 540
 QY 554 AAACACTGATTCAAAATTTTGACCTATTCTCTCTGCTTCCACATCTCTGGGACTACTGGA 613
 Db 541 AAACAGGGTTCAAAATTTGACCGGTTATGCTTTTCATCATCTCTTGGGACTACTGGA 600
 QY 614 CATATCTCTGTTCTCTTACAGTTCCACCTCTCTTTCAGAGATGTCACATGATGTTTGA 672
 Db 601 CATACCTGGCTCTGACAGTTCCACCTCTCTTTCAGAGATGTCACGTGATGTTGTTA 659

RESULT 9
 BY743788
 LOCUS
 DEFINITION BY743788 RIKEN full-length enriched, bone marrow macrophage Mus
 musculus cDNA clone 1830016H14 5', mRNA sequence.
 ACCESSION BY743788.1 GI:27170046
 VERSION
 KEYWORDS
 SOURCE Mus musculus (house mouse)
 ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 660)

OKazaki,Y., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
 Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
 Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
 Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
 Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
 Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
 Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
 Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
 Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
 Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
 Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
 Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
 Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
 Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
 Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
 Sandelin,A., Schneider,C., Sempile,C.A., Setou,M., Shimada,K.,
 Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
 Verardo,R., Wagner,L., Wahlstedt,C., Wang,Y., Watanabe,Y.,
 Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
 Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
 Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
 Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
 Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
 Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
 Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
 Rogers,J., Birney,E. and Hayashizaki,Y.
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)
 22354683
 12466851
 CONTACT: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Teurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-resgsc.riken.go.jp,
 URL:http://genome.gsc.riken.go.jp/
 Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P.,
 Fukuda,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F.,
 Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y.,
 Kondo,S., Konno,H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M.,

Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N.,
 Sano,H., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M.,
 Takeda,I., Waki,K., Wataniki,A., Muramatsu,M. and Hayashizaki,Y.
 Direct Submission
 Computational Analysis of Full-Length Mouse cDNAs Compared with
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Tissues were provided by David A. Hume (Depts. of Biochemistry
 and Microbiology/Parasitology Institute for Molecular Bioscience
 University of Queensland Brisbane,Q 4072 Australia) whose
 assistance we gratefully acknowledge.
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for
 further details.

Location/Qualifiers
 1. 660

FEATURES
 source

/organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="1830016H14"
 /tissue_type="bone marrow"
 /cell_type="macrophage"
 /clone_lib="RIKEN full-length enriched, bone marrow
 macrophage"

ORIGIN

Query Match 55.6%; Score 506.6; DB 13; Length 660;
 Best Local Similarity 85.9%; Pred. No. 8.2e-136;
 Matches 560; Conservative 0; Mismatches 92; Indels 0; Gaps 0;
 QY 8 CGGCTCGATTCCACCCCGAGGAGCCATGTCGAGGCTGAGTGGGGATACCGGAGCACA 67
 Db 9 CTGCCACAGGCTCTCCATCGAGACCATCGGAGGCTGAGTGGGGTACGGCGAGCACA 68
 QY 68 ACGGTCCTATTCACTGGAAAGGAATTTTCCCTATTGCTGATGGTGATCGCAATCTCCAA 127
 Db 69 ACGGTCCTATTCACTGGATGAATTTTCCCTATTGCTGATGGTGATCGAGTCTCCAA 128
 QY 128 TTGAGATTAAACCAAGAGTGAATATGACTCTTCCCTCCGACCACTAGTATCAAGT 187
 Db 129 TTGAGATTAAACCAAGAGTGAATATGACTCTTCCCTCCGACCACTAGTATCAAGT 188
 QY 188 ATGACCCAGCTCAGCTAAATCATCAGCAACAGCGGCCATTCCTCAATGTTGACTTTG 247
 Db 189 ATGATCTCTCCCTAGCTAAATCATCAGCAATAGTGGCCATTCCTTCAAGTTGACTTTG 248
 QY 248 ATGACACAGAGAACAAATCAGTTTCTGCGTGGTGGTCTCTCACTGGAAGCTACAGTTTAC 307
 Db 249 ACGACACGAGGACAAATCAGTTTCTGCGTGGAGGTCTCTCACTGGGAACTACAGGTTGC 308
 QY 308 GCGAGTTTCACTTCACTGGGGTCCGCTGATGACCAACCGCTCCGAGCACAATAGTAGATG 367
 Db 309 GCGAGTTTCACTTGGCACTGGGGTTCAGACAGATGATCATCGCTCAGAGCATGTTGATAGC 368
 QY 368 GAGTGAGCTATGCTGCAGAGCTCCCATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 427
 Db 369 GAGTGAGCTATGCTGCAGAGCTCCCATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 428
 QY 428 TTGTTGAGCAGCTCATGAACACAGATGGAATCGGCTGCTTTGGAGAGTGTGTTTACAGATTG 487

Db 429 TTGTAGAGGCGAGCTCATGAGTCCGATGGCTGGCTGTCTCTGGAGTATTTTCTACAGATTG 488
 QY 488 GTGAACCTTAATTCCTCAACTGCAAAAGATTAATGACACATTTTGATTCATTAAGAAAAGG 547
 Db 489 GGGAAACACAATCTCTCACTGACANAAGATCACTGATATTTTGGATTCCATANAGGANAAGG 548
 QY 548 GTAACAAACTCGATTGACAAATTTTGACCTATTGCTCTGCTCCACATCCCTGGGACT 607
 Db 549 GTAAACAAACACGGTTTCAAAATTTTGACCCGGTTATGCTGCTCCATCATCTTGGGACT 608
 QY 608 ACTGGACATATCTCTGTTCTCTTACAGTTCCACCTCTTCTTGGAGAGTGTAC 659
 Db 609 ACTGGACATACCTGGCTCTCTGACAGTTCCACCTCTTCTTGGAGAGTGTAC 660

RESULT 10
 BY744810 666 bp mRNA linear EST 17-DEC-2002
 DEFINITION BY744810 RIKEN full-length enriched, bone marrow macrophage Mus
 musculus cDNA clone I830048H05 5', mRNA sequence.

ACCESSION BY744810
 VERSION BY744810.1 GI:27171671

KEYWORDS EST.
 SOURCE Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 666)

REFERENCE
 AUTHORS Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
 Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I.,
 Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
 Hume, D.A., Quackenbush, J., Baldarelli, R., Hill, D.P., Bult, C.,
 Batalov, S., Beisel, K.W., Blake, J.A., Bradt, P., Brusci, V.,
 Chotia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,
 Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,
 Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,
 Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,
 Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,
 Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,
 Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
 Numata, K., Okido, T., Pavan, W.J., Perte, G., Pesole, G.,
 Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S.,
 Ravasi, T., Reed, J.C., Reid, J., Reid, J., Ring, B.Z., Ringwald, M.,
 Sandelin, A., Schneider, C., Sempile, C.A., Setou, M., Shimada, K.,
 Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,
 Vextardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,
 Wells, C., Wilming, L.G., Wyszewski-Boris, A., Yanagisawa, M., Yang, I.,
 Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,
 Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,
 Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
 Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,
 Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,
 Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,
 Rogers, J., Birney, E. and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216

Email: genome-res@sc.riken.go.jp
 URL: http://genome.gsc.riken.go.jp/
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Hirozane, T., Hori, F.,
 Fukuda, S., Hashizume, W., Hayashida, K., Hashizume, W., Hori, F.,
 Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y.,
 Kondo, S., Konno, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M.,

FEATURES

source

Location/Qualifiers

1..666
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="I830048H05"
 /tissue_type="bone marrow"
 /cell_type="macrophage"
 /clone_lib="RIKEN full-length enriched, bone marrow
 macrophage"

ORIGIN

Query Match 55.2%; Score 502.6; DB 13; Length 666;
 Best Local Similarity 85.8%; Pred. No. 1.2e-134;
 Matches 567; Conservative 0; Mismatches 93; Indels 1; Gaps 1;
 QY 8 CGGCTCGAGTTCCACCCGAGGACCATGTGAGGCTCAGCTGGGATACCGGAGCACA 67
 Db 7 CTGCCACAGGCTCTCCATCGAGGACCATGGCGAGGCTGAGCTGGGGTACGCGAGCACA 66
 QY 68 ACGGCTCTATTACATGGGAAGGAATTTTCCCTATTGCTGATGCTGATCAGCAATCTCCAA 127
 Db 67 ACGTCCCATTCATCGGAATGNAATTTGCCCTATTGCTGATGCTGATCAGCAATCTCCAA 126
 QY 128 TTGAGATTAAACCAAGAGTGAATATGACTTTCCTCCGACCACTTAGTATCAAGT 187
 Db 127 TTGAGATTAAACCAAGAGTGAATATGACTTTCCTCCGACCACTTAGTATCAAGT 186
 QY 188 ATGACCCAGCTCAGCTAAATCATCAGCAACAGCGGCATTCCTCAATGTTGACTTG 247
 Db 187 ATGATCCTCGCTCAGCTAAATCATCAGCAATAGTGGCCATTCCTTCAACGTTGACTTG 246
 QY 248 ATGACACAGAGAAACAAATCAGTTCTCGTGGTGGTCTCTCTCATCTGGAAGCTACAGTTAC 307
 Db 247 ACACACGAGGACAAATCATGTTCTCGTGGAGGTCCTCTCATCTGGAAGCTACAGTTAC 306
 QY 308 GGCAGGTTTCACTTCCACTGGGGTCCCGCTGATGACCAACCGGCTCCGAGCACAATAGTAGT 367
 Db 307 GGCAGTTCCATTGCACTGGGGTCCCGCTGATGATGATGATGATGATGATGATGATGATG 365
 QY 368 GAGTGAAGTATGCTCGAGCTCCATGTTGTTTCACTGGAAATTCAGACAAATACCCAGCT 427
 Db 366 GAGTGAAGTATGCTCGAGCTCCATGTTGTTTCACTGGAAATTCAGACAAATACCCAGCT 425
 QY 428 TTGTTGAGGAGCTCATGACAGATGAGCTGGCTGCTCTTGGAGTGTGTTTTACAGATTG 487

Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N.,
 Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M.,
 Takeda, Y., Waki, K., Watanabe, A., Muramatsu, M. and Hayashizaki, Y.
 Direct Submission
 Computational Analysis of Full-Length Mouse cDNAs Compared with
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Tissues were provided by David A. Hume (Depts. of Biochemistry
 and Microbiology/Parasitology Institute for Molecular Bioscience
 University of Queensland Brisbane, Q 4072 Australia) whose
 assistance we gratefully acknowledge.
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for
 further details.


```

Db      426 TTGTAGAGGCGAGCTCATGAGTCGATGGCTGGCTGCTGNGAGTATTTCTACAGATTG 485
QY      488 GTGACCTTAATTCCTCCAACTGCGAAAGATTACTGACACTTTGGATTCCATTAAAGAAAAGG 547
Db      486 GGGAAACACAATCCTCAACTGCGAAAGATCACTGATATTTTGGATTCCATANAGGAAAAGG 545
QY      548 GTAACAAACTCGATTACCAAAATTTTGACCTATTGCTCTGCTGCTCCACCATCTGGGACT 607
Db      546 GTAACAAACACGGTTCACAAATTTGACCCGTTATGCTTGTCTTCATCATCTTGGGACT 605
QY      608 ACTGGACATATCCTGGTCTCTTACAGTTCCACCTCTTCTTGAGAGTGTACATGGAATTG 667
Db      606 ACTGGACATACCTGCTCTCTGACAGNCCACCTCTTCTTGAGAGTGTACATGGAATTG 665
QY      668 T 668
Db      666 T 666

BY744678      656 bp      mRNA      linear      EST 17-DEC-2002
BY744678      RIKEN full-length enriched, bone marrow macrophage Mus
musculus cDNA clone I830044H06 5', mRNA sequence.
BY744678      1      GI:27171468
EST.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 656)
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I.,
Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C.,
Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,
Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusci, V.,
Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,
Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,
Gustinich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,
Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,
Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,
Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
Nunata, K., Okido, T., Pavan, W.J., Perte, G., Pesole, G.,
Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S.,
Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M.,
Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K.,
Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,
Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,
Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,
Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,
Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,
Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,
Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,
Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,
Rogers, J., Birney, E. and Hayashizaki, Y.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
12466851
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp
URL: http://genome.gsc.riken.go.jp/

```

```

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P.,
Fukuda, S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F.,
Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y.,
Kondo, S., Konno, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M.,
Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N.,
Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M.,
Takeda, Y., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken.
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Tissues were provided by David A. Hume ( Depts. of Biochemistry
and Microbiology/Parasitology Institute for Molecular Bioscience
University of Queensland Brisbane, Q 4072 Australia ) whose
assistance we gratefully acknowledge.
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
FEATURES
Location/Qualifiers
1. 656
/organism="Mus musculus"
/mol_type="mRNA"
/strain="CS7BL/6J"
/db_xref="taxon:10090"
/clone="I830044H06"
/tissue_type="bone marrow"
/cell_type="macrophage"
/clone_lib="RIKEN full-length enriched, bone marrow
macrophage"
ORIGIN
Query Match 53.7%; Score 489.6; DB 13; Length 656;
Best Local Similarity 85.4%; Pred. No. 7.2e-131;
Matches 554; Conservative 0; Mismatches 94; Indels 1; Gaps 1;
QY 8 CGGCTCGAGTTCCACCCCGAGGAGCCATGTCGAGGCTCAGCTGGGGATACCGGAGCACA 67
Db 9 CTGCCACAGGCTCTCCATCGAGGACCATGGCGAGGCTGAGCTGGGGGTACGGCGAGCACA 68
QY 68 AGGTCCTTATTCAGCTGGAGGATTTTCCCTATTGCTGATGCTGATCAGCAATCTCCAA 127
Db 69 ACGGTCCTCAATCAGTGGAAATGAATGTTCCCTATTGCTGATGCTGATCAGCTCTCCAA 128
QY 128 TTGAGATTAAACCAAGAGTGAATATGACTCTTCCCTCCGACCACCTAGTATCAAGT 187
Db 129 TTGAGATTAAACCAAGAGTGAATATGACTCTTCCCTCCGACCACCTAGTATCAAGT 188
QY 188 ATGACCCCAAGCTCAGCTAAATATCATGACCAACAGCGGGCCATTCCTTCAATGTTGATTG 247
Db 189 ATGATCCTGCTCAGCTAAATATCATGACCAATAGTGGCCATTCCTTCAACGTTGACTTTG 248
QY 248 ATGACACAGAGCAACAAATCAGTTCTGCGTGGTGGTCTCTCTCACTGGAAGCTCAGGTTAC 307
Db 249 ACGACACGAGGACAAATCAGTTCTGCGTGGAGGTTCTCTCTCACTGGAACTCAGGTTGC 308
QY 308 GGCAGGTTCACTTCACTGGGGGTCCGCTGATGACCCAGCGGCTCCGAGCACAATAGTAGATG 367
Db 309 GGCAGTTTCATTGCACTGGGGGTCCAGCAGATGATCATGGCTCAGAGCATGTGGTAGACG 368
QY 368 GAGTGAAGTATGCTGAGAGCTCCCAATGTTGTTCACTGGAATTCAGCAAAATACCCAGCT 427

```

```

Db      369 GAGTGAGGTGCTCCAGAGCTGCATGTTGTCCTCAGTGAATTCAGACAAATACCCGAGCT 428
Qy      428 TTGTTGAGCGAGCTCATGAACAGATGAGTGGCTGTCTTGGAGTGTTTTTACAGATTG 487
Db      429 TTGTAGAGCGAGCTCATGAGTCGATGGCTGGCTGTCTCTGNGAGTATTTTTCACAGATTG 488
Qy      488 GTGAACCTTAATTCCTCAACTGCAAGATTAATGACACTTTGGATTCCATTAAAGAAAGG 547
Db      489 GGGAAACAATCTCTCAACTGCAAGATTAATGACACTTTGGATTCCCA-TAAAGAAAGG 547
Qy      548 GTAAACAAACTCGATTCAAAATTTTGACCTATTGCTCTGCTTCCACCATCTCTGGAGCT 607
Db      548 GTAAACAAACAGGTTCAAAATTTTGACCTTTTACCCGTTATGCTTTTCAATCACTCTGAGCT 607
Qy      608 ACTGGACATATCTCGTTCTCTTACAGTTCCACCTCTTCTTTCAGAGTGT 656
Db      608 ACTGGACATATCTCTCTCTGACAGTCCACCTCTTCTTCTGAGAGTGT 656

RESULT 12
CF768871
LOCUS   CF768871
DEFINITION Bos taurus skin cDNA library Bos taurus cdna clone
ACCESSION CF768871
VERSION  CF768871.1 GI:37718090
KEYWORDS Bos taurus (cow)
SOURCE  Bos taurus
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 692)
AUTHORS Wang, Y.H., McWilliam, S. and Lehnert, S.
TITLE Transcription profiling of cattle skin
JOURNAL Unpublished (2003)
COMMENT Contact: Dr Yonghong Wang
Functional Genomics Lab
CSIRO Livestock Industries
Level 5, Queensland Biosciences Precinct, University of Queensland,
306 Carmody Road St. Lucia QLD Australia
Tel: 07 3214 2445
Fax: 07 3214 2685
Email: Yonghong.Wang@csiro.au
Plate: 2 row: G column: 4.
FEATURES             source
1..692
/organism="Bos taurus"
/mol_type="mRNA"
/strain="Hereford Shorthorn"
/db_xref="taxon:9913"
/clone="CCL000987"
/sex="female"
/tissue_type="pooled"
/dev_stage="Adult"
/lab_host="XLI-BlueMRF'strain"
/clone_lib="Bos taurus skin cdna library"
/note="Organ: skin; Vector: Uni-ZAPXR; Site 1: EcoRI;
Site 2: Xho I; Library made from pooled skin of adult
female Hereford-Shorthorn."

ORIGIN
Query Match 52.5%; Score 478; DB 14; Length 692;
Best Local Similarity 87.1%; Pred. No. 1.8e-127;
Matches 559; Conservative 0; Mismatches 80; Indels 3; Gaps 3;

Qy      53 GATACGGCAGACACAGCTCTTATCTACCTGAAGAAATTTTCCCTATTGCTGATGGTG 112
Db      1 GCTACTGCGCGCGCAGCTCTCTGTTCATTGGAATGAATTTTCCCTATTGCTGATGGGG 60
Qy      113 ATCAGCAATCTCCAATTGAGATTAAACCAAGAAAGTGAATATGACTCTTCCCTCCGAC 172
Db      61 ATCAACAATCTCGATTGAGATTAAACCAAGAAAGTGAATATGATTCTTCCCTCCGAC 120

```

```

Qy      173 CACTTAGTATCAAGTATGACCCAGCTCAGCTAAATCATCAGCAACGCGCCATTCTT 232
Db      121 CTCCTGGTATCAAGTATGACCTCAGCTCAGCTAAATCATCAGTAAATAGTGCCACTCT 180
Qy      233 TCAATGTTGACTTTTATGACACAGAGAAATCAAGTTCTGGTGGTGGTCTCTCTCAGCTG 292
Db      181 TCAAGTTGACTTTTATGACACAGAGAAATCAAGTTCTGGTGGTGGTGGTGGTGGTGGTGG 240
Qy      293 GAAGCTACAGGTTACGGCAGGTTTCACTTCACTGGGGGTCCGCTGATGACCAAGGCTCGG 352
Db      241 GAAGCTACAGGTTACGGCAGGTTTCACTTCACTGGGGGTCCGCTGATGATGATGATGATG 300
Qy      353 AGCATATAGTAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 412
Db      301 AACATGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 360
Qy      413 ACAAAATACCCAGCTTTTGTGAGGAGCTTCATGAACACAGATGAGTGGCTGTCTTGGGAG 472
Db      361 ACAAAATACCCAGCTTTTGTGAGGAGCTTCATGAACACAGATGAGTGGCTGTCTTGGGAG 420
Qy      473 TGTTTTACAGATTGGTGAACCTAATCCCACTGCAAAAGATTACTGACACTTTGGATT 532
Db      421 TATTTTACAGATTGGTGAACATAATCCCACTGCAAAAGATTACTGACACTTTGGATT 480
Qy      533 CCATTAAGAAAGAGGTAAACAACTCGATTCAAAATTTTGCACCTATGCTCTGCTTC 592
Db      481 CCATTAAGAAAGAGGTAAACAACTCGATTCAAAATTTTGCACCTATGCTCTGCTTC 540
Qy      593 CACCATCTGGGACTACCTGGACATA-TCTGTGTTCTTACAGTTCACCTC-TTCTTGA 650
Db      541 CTCATGCGGGATTATTGGACATATTCTGCTCTCTTACGGTCCACCTCTTCTTCTGGA 600
Qy      651 GAGTGTACAC-ATGGATTGTTTAAAGCAACCTTATAAATCA 691
Db      601 AAGCGTCACTTTGGATCATTTAAAGCAATCCATAAATCA 642

RESULT 13
BY722663
LOCUS   BY722663
DEFINITION RIKEN full-length enriched, adult male bone Mus musculus
ACCESSION BY722663
VERSION  BY722663.1 GI:27135780
KEYWORDS Mus musculus (house mouse)
SOURCE  Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 639)
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
Nikaido, I., Osato, N., Saito, R., Saito, H., Yamaguchi, Y., Nogami, A.,
Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Kanapin, A.,
Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C.,
Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,
Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusci, V.,
Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,
Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,
Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,
Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,
Kawaji, H., Kawasaki, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,
Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,
Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G.,
Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S.,
Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M.,
Sandelin, A., Schneider, C., Sempole, C.A., Setou, M., Shimada, K.,
Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,
Verrardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,
Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,
Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,
Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,

```

modified pBluescript
FLC I."

COMMITMENT

11292378
Contact: Smith TPL
USDA Aps US Meat Animal Research Center

Tel: 402 762 4366
Fax: 402 762 4390
Email: smath@mail.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCAGCTCAGCAGC
Plate: 47 row: F column: 16
Seq primer: ATTAGTGACACTATAG.
Location/Qualifiers
1. 561
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 480V"
/note="Vector: PCMV SPOT6; Site 1: NotI; Site 2: SalI;
Library made from pooled tissue from day 20 and day 40
embryos."

FEATURES
source

ORIGIN

Query Match 49.0%; Score 446.6; DB 10; Length 561;
Best Local Similarity 88.3%; Pred. No. 2.2e-118; Indels 1; Gaps 1;
Matches 496; Conservative 0; Mismatches 65

QY 47 GCTGGGGTACCGGAGCACAACGGTCTCTATTCACTGGAAGAAATTTTCCCTATTGCTG 106
DB 1 GCTGGGGTACCGGAGCAGCGGCTCTGTTCAATGGAATGAATTTTCCCTATTGCTG 60

QY 107 ATGGTGATCAGCAATCTCCAATGAGATTAACCAAGAGTGAATATGACTCTTCCC 166
DB 61 ATGGGGATCAACAATCTCCGATGAGATTAACCAAGAGTGAATATGACTCTTCCC 120

QY 167 TCCGACCACTTAGTATCAAGATGATGCCAAGCTCAGTAAATCATCAGCAACAGCGGC 226
DB 121 TCCGACCTCTTGATCAAGATGAGTGCAGCTCAGCTAAATCATCAGTAATAGTGGCC 180

QY 227 ATTCTTCAATGTGACTTTGATGACACAGAGAAACAAATCAGTTCTGCGTGGTGGTCTC 286
DB 181 ACTCTTCAACGTGACTTTGACGACACGAGCAAAATCAGTTCTGCGTGGGCCCCC 240

QY 287 TCAGTGAAGCTACAGGTTACGGAGGTTCACTTCACTGGGGTCCGCTGATGACACAG 346
DB 241 TCAGTGAAGCTACAGGTTACGGAGTTCACCTTCACTGGGGTCCACTGATGATCAG 300

QY 347 GCTCCGAGCACATAGTAGTGAGTGAGCTATGCTGCAGAGCTCCATGTTGTTCACTGA 406
DB 301 GCTCTGAACATGTTGGTGGATGGGTAGGTACGCTGCAGAGCTCCACGTTGTTCACTGA 360

QY 407 ATTGAGCAAAATACCCAGCTTTGTTGAGCGAGCTCATGAACAGATGAGCTGGCTGCT 466
DB 361 ATTGAGCAAAATACCCAGCTTTGTTGAGCGAGCTCATGAGCGAGATGAGCTAGCTGT 420

QY 467 TGGGAGTGTTTTACAGATGTTGAACTTAATTCCTCACTGCAAAAGATTAAGTACACTT 526
DB 421 TGGGAAT-ATTTTACAGATGTTGTAACATAATCCCACTGCAAAAGATTAAGTACACTT 479

QY 527 TGATTTCCATTAAAGAAAAGGTAAACAACTCGATTCAAAATTTTGACCTATTGCTC 586
DB 480 TGATTTCCATTAAAGAAAAGGTAAACAACTCGATTCAAAATTTTGACCTATTGCTC 539

QY 587 TGCTTCCACCTCCTGGGACTA 608
DB 540 TACTTCTCCATGCGGGATTA 561

RESULT 15
BU169155 928 bp mRNA linear EST 04-SEP-2002
LOCUS BU169155
DEFINITION AGENCOURT_7974620 NIH_MGC_110 Homo sapiens cdna clone IMAGE:6083105

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BU169155
BU169155.1 GI:22683139
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2312 row: 1 column: 18
High quality sequence stop: 539.

FEATURES

source

1. 928
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6083105"
/tissue_type="ductal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_110"
/note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; CDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

ORIGIN

Query Match 47.8%; Score 435.8; DB 13; Length 928;
Best Local Similarity 94.6%; Pred. No. 3.8e-115; Indels 5; Gaps 2;
Matches 474; Conservative 0; Mismatches 22

QY 17 TTCACCCCGAGGACCATGTCGAGGCTCAGCTGGGGATACCGGAGCAACACGGTCTTA 76
DB 285 TTCACCCCGAGGACCATGTCGAGGCTCAGCTGGGGATACCGGAGCAACACGGTCTTA 344

QY 77 TTCACCTGGAGAAATTTTCCCTATTGCTGATGGTATCAGCAATCTCCAATTGAGATTA 136
DB 345 TTCACCTGGAGAAATTTTCCCTATTGCTGATGGTATCAGCAATCTCCAATTGAGATTA 404

QY 137 AACCAAAAGAACTGAATATGACTCTCCCTCCGACCACTTAGTATCAAGTATGACCA 196
DB 405 AACCAAAAGAACTGAATATGACTCTCCCTCCGACCACTTAGTATCAAGTATGACCA 464

QY 197 GCTCAGCTAAAATCATCAGCAACAGCGGCATTCCTTCAATGTTGACTTTGATGACACAG 256
DB 465 GCTCAGCTAAAATCATCAGCAACAGCGGCATTCCTTCAATGTTGACTTTGATGACACAG 524

QY 257 AGAAACAAATCAGTTCTGGTGGTGTCTCTCTACCTGGAAGCTACAGGTTACGGCAGGTTTC 316
DB 525 AGAAACAAATCAGTTCTGGTGGTGTCTCTCTACCTGGAAGCTACAGGTTACGGCAGGTTTC 584

QY 317 ACCTTCACCTGGGGTCCGCTGATGATGACCGGCTCCGAGCACATAGTAGTACGAGT 376
DB 585 ACCTTCACCTGGGGTCCGCTGATGATGACCGGCTCCGAGCACATAGTAGTACGAGT 644

QY 377 ATGCTGCAGAGCTCCATGTTGTTCACTGGAATTCAGACAAATACCCAGCTTTGTTGAGG 436
DB 645 ATGCTGCAGAGCTCCCTGTTGTTCACTGGAATTCAGACAAATACCCAGCTTTGTTGAGG 704

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 2004, 22:29:06 ; Search time 96 Seconds
(without alignments)
5266.253 Million cell updates/sec

Title: US-10-069-434-4

Perfect score: 911
Sequence: I CGGAATCGGTCGAGTCC.....CCCCAATCCAGGGGAAAA 911

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCUTS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	284.8	31.3	605	1	US-08-469-667-17
2	284.8	31.3	605	4	US-09-224-110-17
3	284.8	31.3	605	5	PCR-US95-07289-17
4	280	30.7	780	3	US-08-927-128-3
5	280	30.7	864	2	US-08-595-868C-11
6	280	30.7	867	3	US-09-139-819A-11
7	280	30.7	867	4	US-09-750-913-11
8	280	30.7	1759	4	US-09-576-594-643
9	278.4	30.6	1755	4	US-09-566-921-124
10	266.8	29.3	1266	4	US-09-938-270B-2
11	177.4	19.5	599	3	US-09-385-982-446
12	147.6	16.2	477	4	US-09-621-976-18059
13	147.4	16.2	581	3	US-09-385-982-136
14	136.8	15.0	760	3	US-09-385-982-156
15	111	12.2	586	3	US-09-385-982-472
16	92.6	10.2	55298	4	US-09-491-356C-1
17	88.2	9.7	2721	3	US-08-921-195-1
18	88	9.7	45716	4	US-08-965-048-5
19	88	9.7	45989	4	US-08-965-048-6
20	87.8	9.6	35060	3	US-08-814-095-7
21	86.8	9.5	66804	4	US-09-740-041-3
22	86.6	9.5	1624	2	US-08-852-807-10
23	86.6	9.5	13674	2	US-08-852-807-1
24	86.6	9.5	39982	4	US-09-820-924-3
25	86.6	9.5	161652	4	US-09-497-855A-40
26	86	9.4	392000	4	US-10-027-983-11
27	85.8	9.4	18073	3	US-09-078-294-12
17					Sequence 17, Appl
17					Sequence 17, Appl
17					Sequence 17, Appl
3					Sequence 3, Appl
11					Sequence 11, Appl
11					Sequence 11, Appl
11					Sequence 11, Appl
643					Sequence 643, Appl
2					Sequence 2, Appl
446					Sequence 446, Appl
18059					Sequence 18059, A
136					Sequence 136, Appl
156					Sequence 156, Appl
472					Sequence 472, Appl
1					Sequence 1, Appl
5					Sequence 5, Appl
6					Sequence 6, Appl
7					Sequence 7, Appl
3					Sequence 3, Appl
10					Sequence 10, Appl
1					Sequence 1, Appl
3					Sequence 3, Appl
40					Sequence 40, Appl
11					Sequence 11, Appl
12					Sequence 12, Appl

Sequence 8244, Ap
Sequence 8253, Ap
Sequence 12925, A
Sequence 35, Appl
Sequence 2, Appl
Sequence 3, Appl
Sequence 2, Appl
Sequence 3, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 2, Appl
Sequence 9, Appl
Sequence 1, Appl
Sequence 5, Appl
Sequence 2, Appl

US-09-621-976-8244
US-09-621-976-8253
US-09-621-976-12925
US-09-103-663-35
US-09-033-333-3
US-09-033-333-2
US-09-614-495-3
US-09-151-376-2
US-08-380-916-1
US-08-721-690-1
US-08-891-581-1
US-09-033-333-2
US-09-033-556-1
US-09-614-495-2
US-09-474-699-9
US-09-151-376-1
US-09-318-448-5
US-09-345-217-2

ALIGNMENTS

RESULT 1
US-08-469-667-17
; Sequence 17, Application US/08469667
; Patent No. 5733748
; GENERAL INFORMATION:
; APPLICANT: Yu, Guo-Liang
; TITLE OF INVENTION: Colon Specific Genes and Proteins
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi,
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: NJ
; COUNTRY: USA
; ZIP: 07068-1739
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,667
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferraro, Gregory D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-435
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 605 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..603
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 1..603
US-08-469-667-17

Query Match 31.3%; Score 284.8; DB 1; Length 605;
Best Local Similarity 67.2%; Pred. No. 3.8e-84;


```
QY 48 CTGGGATACCGGAGCAACAGGTCCTATTCTAGGAGGATTTTCCCTATTCTCTGA 107
Db 12 CTGGGGGTACGGCAACACAGGACCTGAGCACTGGCATAGGACTTCCCAATGCCAA 71
QY 108 TGTGATCAGCAATCTCCAAATTCAGATTAAACCAAGAAAGTGAAATATGACTTCCCT 167
Db 72 GGGAGAGCGCCAGTCCCTGTTGACATCGACACTCATACAGCAAGATGACCCCTCCCT 131
QY 168 CCGACCACTTAGTATCAAGTATGACCAAGCTCAGCTAAATCATCAGCAACAGCGCCA 227
Db 132 GAAGCCCTGTCTGTTTCCTATGATCAAGCAACTTCCCTGAGGATCTCAACAATGGTCA 191
QY 228 TTCTTTCAATGTTGACTTTGATGACACAGAGACAAATCAGTTCTCGTGGTCTCT 287
Db 192 TGTCTTCAAGTGGAGTTTGGATGACTCTCAGGCAAGAGAGTGCTCAAGGGAGGACCCCT 251
QY 288 CACTGGAAGCTACAGGTTACGGCAGGTTCACTTCACTGGGGTCCGCTGATGACCACGG 347
Db 252 GGATGSCACTTACAGATTGATTCAGTTTCACTTCACTGGGTTCACTTGGTGGACAAG 311
QY 348 CTCGGAGCAGATAGTAGAGTGAGCTATGCTGAGAGCTCCATGTTGTTCACTGGAA 407
Db 312 TTTCAGAGCATACTGTGATAAAGAAATATGCTGCAGAACTTCACTTGTCTCACTGGAA 371
QY 408 TTCAGACAAATACCCAGCTTTGTTGAGGAGCTCATGAACAGATGAGTGGCTGTCT 467
Db 372 CAC---CAATATGGGATTTGGGAAGCTGTGACAGCAAGAGTGCTCAAGGGAGGACCCCT 428
QY 468 GGGAGTGTGTTTACAGATTGGTGAACCTTAATTTCCCACTGCAAAAGATTTACTGACACTTT 527
Db 429 AGGTATTTTGAAGTTGGCAGCGCTAAACCGGSCCTTCAGAAAGTTGTTGATGTCT 488
QY 528 GGAATCCATTAAGAAAGGGTAAACAACTGATTAACAAATTTGACCTATGTCCT 587
Db 489 GGAATCCATTAAGAAAGGGTAAAGAGTGCTGACTTCACTTAATTCGATCCTCGTGGCT 548
QY 588 GCTTCCACCATCTGGGACTGAGTACATATCTGTTCTCTTACAGTTCCACCTCTCT 647
Db 549 CTTTCTGATCTCTGATTTGAGTCTGAGCTTACCAGGCTCACTGACCAACCCCTCTCT 608
QY 648 TGAGAGTGTCATGATTTGTTTAAAGCAACCTATAAATACATCAGCTCTCAACAGTGGC 707
Db 609 GGAATGTCGACCTGATTTGCTCAAGGAACCCATCAGCTCAGCAGCAGCAGGTT 668
QY 708 CAAATTCGAGTCTCTGTCAGCAGCGAGGTTGAAGCAGAG 751
Db 669 GAAATTCGTAACCTTAACTTCAATGGGAGGTTGAACCCGAAG 712
```

RESULT 5

```
US-08-595-868C-11
; Sequence 11, Application US/08595868C
; Patent No. 5962270
; GENERAL INFORMATION:
; APPLICANT: Wagner, Fred
; APPLICANT: Stout, Jay
; APPLICANT: Henriksen, Dennis
; APPLICANT: Partridge, Bruce
; APPLICANT: Holmquist, Bart
; APPLICANT: Frank, Julie
; TITLE OF INVENTION: RECOMBINANT PREPARATION OF CALCITONIN FRAGMENTS AND US
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant & Gould
; STREET: 3100 No. 5962270west Center, 90 S. 7th Street
; CITY: Minneapolis
; STATE: MN
; COUNTRY: U.S.A.
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
```

```
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/595,868C
; FILING DATE: 06-FEB-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Carter, Charles G
; REGISTRATION NUMBER: 35,093
; REFERENCE/DOCKET NUMBER: 8648.59US01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612/332-5300
; TELEFAX: 612/332-9081
; TELEX:
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 864 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1...864
; OTHER INFORMATION:
; US-08-595-868C-11
```

```
Query Match 30.7%; Score 280; DB 2; Length 864;
Best Local Similarity 63.4%; Pred. No. 1.8e-82;
Matches 446; Conservative 0; Mismatches 255; Indels 3; Gaps 1;
```

```
QY 48 CTGGGATACCGGAGCAACAGGTCCTATTCTAGGAGGATTTTCCCTATTCTCTGA 107
Db 12 CTGGGGGTACGGCAACACAGGACCTGAGCACTGGCATAGGACTTCCCAATGCCAA 71
QY 108 TGTGATCAGCAATCTCCAAATTCAGATTAAACCAAGAAAGTGAAATATGACTTCCCT 167
Db 72 GGGAGAGCGCCAGTCCCTGTTGACATCGACACTCATACAGCAAGATGACCCCTCCCT 131
QY 168 CCGACCACTTAGTATCAAGTATGACCCAGCTCAGCTAAATCATCAGCAACAGCGCCA 227
Db 132 GAAGCCCTGTCTGTTTCCTATGATCAAGCAACTTCCCTGAGGATCCTCAACAATGGTCA 191
QY 228 TTCTTTCAATGTTGACTTTGATGACACAGAGACAAATCAGTTCTCGTGGTGGTCTCT 287
Db 192 TGTCTTCAAGTGGAGTTTGGATGACTCTCAGGACAAAGCAGTGTCTCAAGGGAGGACCCCT 251
QY 288 CACTGGAAGCTACAGGTTACGGCAGGTTCACTTCACTGGGGTCCGCTGATGACCACCG 347
Db 252 GGATGSCACTTACAGATTGATTCAGTTTCACTTCACTGGGTTCACTTGGTGGACAAG 311
QY 348 CTCGGAGCAGATAGTAGAGTGAGCTATGCTGAGAGCTCCATGTTGTTCACTGGAA 407
Db 312 TTTCAGAGCATACTGTGATAAAGAAATATGCTGCAGAACTTCACTTGTCTCACTGGAA 371
QY 408 TTCAGACAAATACCCAGCTTTGTTGAGGAGCTCATGAACAGATGAGTGGCTGTCT 467
Db 372 CAC---CAATATGGGATTTTGGGAAGCTGTGAGCAACCTGATGAGTGGCTGGCTCT 428
QY 468 GGGAGTGTGTTTACAGATTGGTGAACCTTAATTTCCCACTGCAAAAGATTTACTGACACTTT 527
Db 429 AGGTATTTTGAAGTTGGCAGCGCTAAACCGGSCCTTCAGAAAGTTGTTGATGTCT 488
QY 528 GGAATCCATTAAGAAAGGGTAAACAACTGATTAACAAATTTGACCTATGTCCT 587
Db 489 GGAATCCATTAAGAAAGGGTAAAGAGTGCTGACTTCACTTAATTCGATCCTCGTGGCT 548
```


;;
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/750,913
;; FILING DATE: 12-Jan-2001
;; CLASSIFICATION: <Unknown>
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 09/139,819
;; FILING DATE: <Unknown>
;; ATTORNEY/AGENT INFORMATION:
;; NAME: BENT, Stephen A.
;; REGISTRATION NUMBER: 29,768
;; REFERENCE/DOCKET NUMBER: 089187/0144
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202) 672-5300
;; TELEFAX: (202) 672-5399
;; INFORMATION FOR SEQ ID NO: 11:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 867 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 1..864
;; SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-09-750-913-11

Query Match 30.7%; Score 280; DB 4; Length 867;
Best Local Similarity 63.4%; Pred. No. 1.8e-82;
Matches 446; Conservative 0; Mismatches 255; Indels 3; Gaps 1;

QY	48	CTGGGATACCGGACCAACAGGTCCTATTCTACTGGAAGGATTTTCCCTATTCCTGA	107
Db	12	CTGGGGGTACGGCAACACCAAGGACCTGAGCACTGGCATAGGACTTCCCCATTGCCAA	71
QY	108	TGCTGATCATGCAATCTCCAAATGAGATTAAACCAAGAAAGTGAATATGACTTCTCCCT	167
Db	72	GGGAGAGCCGACGTCCCTGTTGACATCGACACTCATACGCCAAGTATGACCTTCCCT	131
QY	168	CGGACCACTTAGTATCAAGTATGACCCAGCTCAGCTAAATCATCAGCAACAGCGGCA	227
Db	132	GAAGCCCTGTCTGTTTCTATGATCAAGCACTTCCCTGAGGATCCTCAACAAATGGTCA	191
QY	228	TTCTTCAATGTTGACTTTTGATGACACAGAGAAACAAATCAGTTCTCGGTGGTCTCT	287
Db	192	TGCTTTCACGTGGAGTTTGATGACTCTCAGHCAAGAGTGCTCAAGGAGGACCCCT	251
QY	288	CACGGAAGCTACAGTTACGGCAGGTTCACTTCTACTGGGGTCCGCTGATGACCACGG	347
Db	252	GGATGGCACTTACAGATTGATTCAGTTTCAGTTTCACTGGGTTCACTTGATGGACAAG	311
QY	348	CTCCGAGCACATAGTAGAGTGGAGTACGTATGCTGACAGAGTCCATGTTTTCACATGGA	407
Db	312	TTCAGAGCACTACTGTGGATTAAGAAAGAAATATGCTGCAGAACTTCACTTGGTTCACT	371
QY	408	TTCAGACAAATACCCAGCTTTGTTGAGGAGCTCATGAACAGAGTGGAGTGGCTCTCT	467
Db	372	CAC---CAATATGGGGATTTGGGAAAGCTGTGACAGACCTGATGGAGTGGCCGTTCT	428
QY	468	GGGAGTGTGTTTACAGATTGGTGAACCTTAATTCCTCCAACTGCAAAAGATTACTGACAT	527
Db	429	AGGTATTTTTTGAAGTTTGGCAGCGCTAAACCGGGCCCTTCAGAAAAGTTGTTGATGCT	488
QY	528	GGATTCATTAAGAAAGGTAACAAACTCGATTCAACAATTTGACCTATTGTCTCT	587
Db	489	GGATTCATTAAGAAAGGCAAGAGTGTGATCTTCACTAACTTCGATCCCTCGTGGCCT	548
QY	588	GCTTCCACCATCTGGGACTACTGGACATATCTGTTTCTCTTACAGTTTCCACTTCTCT	647
Db	549	CCTTCTGATCCTTGGATTACTGGACCTACCAGGCTCACTGACACCCCTCTCTTCT	608
QY	648	TGAGATGTCAATGATTTGTTTAAAGCAACCTATAAACATCAGCTCTCAACAGCTGGC	707

Db	609	GGATGTGTGACCTGGATTGTGCTCAAGGAACCCATCAGCTCAGCAGCAGCGGTGT	668
QY	708	CAAATTCGACGTCTCCTGTGCACAGCGGAGGTGAAGCAGCAG	751
Db	669	GAATTCGCTAAACTTAACTTCAATGGGAGGCTGAACCCGAAG	712

RESULT 8

US-09-976-594-643
; Sequence 643, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 643
; LENGTH: 1759
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 2742913CB1
US-09-976-594-643

Query Match 30.7%; Score 280; DB 4; Length 1759;
Best Local Similarity 63.4%; Pred. No. 2.9e-82;
Matches 446; Conservative 0; Mismatches 255; Indels 3; Gaps 1;

QY	48	CTGGGATACCGGACCAACAGGTCCTATTCTACTGGAAGGATTTTCCCTATTCCTGA	107
Db	273	CTGGGGGTACGGCAACACCAAGGACCTGAGCACTGGCATAGGACTTCCCCATTGCCAA	332
QY	108	TGCTGATCATGCAATCTCCAAATGAGATTAAACCAAGAAAGTGAATATGACTTCTCCCT	167
Db	333	GGGAGAGCCGACGTCCCTGTTGACATCGACACTCATACGCCAAGTATGACCTTCCCT	392
QY	168	CCGACCACTTAGTATCAAGTATGACCCAGCTCAGCTAAATCATCAGCAACAGCGGCA	227
Db	393	GAAGCCCTGTCTGTTTCTATGATCAAGCACTTCCCTGAGGATCCTCAACAAATGGTCA	452
QY	228	TTCTTCAATGTTGACTTTTGATGACACAGAGAAACAAATCAGTTCTCGGTGGTCTCT	287
Db	453	TGCTTTCACGTGGAGTTTGATGACTCTCAGGACAAAGCAGTCTCAAGGAGGACCCCT	512
QY	288	CACGGAAGCTACAGTTACGGCAGGTTCACTTCTACTGGGGTCCGCTGATGACCACGG	347
Db	513	GGATGGCACTACTGTGGATTAAGAAAGAAATATGCTGCAGAACTTCACTTGGTTCACT	572
QY	348	CTCCGAGCACATAGTAGAGTGGAGTACGTATGCTGCAGAGTCCATGTTTTCACATGGA	407
Db	573	TTCAGAGCACTACTGTGGATTAAGAAAGAAATATGCTGCAGAACTTCACTTGGTTCACT	632
QY	408	TTCAGACAAATACCCAGCTTTGTTGAGGAGCTCATGAACAGAGTGGAGTGGCTCTCT	467
Db	633	CAC---CAATATGGGATTTTGGGAAAGCTGTGACAGCAACCTGATGGAGTGGCCGTTCT	689
QY	468	GGGAGTGTGTTTACAGATTGGTGAACCTTAATTCCTCCAACTGCAAAAGATTACTGACAT	527
Db	690	AGGTATTTTTTGAAGTTTGGCAGCGCTAAACCGGGCCCTTCAGAAAAGTTGTTGATGCT	749
QY	528	GGATTCATTAAGAAAGGTAACAAACTCGATTCAACAATTTGACCTATTGTCTCT	587
Db	750	GGATTCATTAAGAAAGGCAAGAGTGTGATCTTCACTAACTTCGATCCCTCGTGGCCT	809
QY	588	GCTTCCACCATCTGGGACTACTGGACATATCTGTTTCTCTTACAGTTTCCACTTCTCT	647

Db 810 CTTCTCCTGAATCCTTGGATTACTGGAACCTACCCAGGCTCACTGACCACCCCTCCTCTTCT 869
Qy 648 TCAGAGTGTACATCGATTGTTTAAAGCAACCTATAAACAATCAGCTCTCAACAGCTGGC 707
Db 870 GGAATGTGTGACCTGATTGTCTCAAGAAACCCATCAGCGTCAGCAGCGAGCAGGTGT 929
Qy 708 CAAATTTGCGAGTCTCTGTCGACAGCGGAGGTTGAAGCAGCAG 751
Db 930 GAAATTCGGTAACTTAATTCATTTGGGAGGGTGAACCCGAAG 973

RESULT 9

US-09-566-921-124
; Sequence 124, Application US/09566921
; Patent No. 6682888
; GENERAL INFORMATION:
; APPLICANT: Loring, Jeanne F.
; APPLICANT: Tingley, Debra W.
; APPLICANT: Edwards, Carla M.
; TITLE OF INVENTION: GENES EXPRESSED IN ALZHEIMER'S DISEASE
; FILE REFERENCE: PA-0024 US
; CURRENT APPLICATION NUMBER: US/09/566,921
; CURRENT FILING DATE: 2000-05-05
; NUMBER OF SEQ ID NOS: 138
; SOFTWARE: PERL Program
; SEQ ID NO 124
; LENGTH: 1755
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6682888 248306.1
; NAME/KEY: unsure
; LOCATION: 1751-1752
; OTHER INFORMATION: a, t, c, g, or other
US-09-566-921-124

Query Match 30.6%; Score 278.4; DB 4; Length 1755;
Best Local Similarity 53.2%; Pred. No. 9.7e-82;
Matches 445; Conservative 0; Mismatches 256; Indels 3; Gaps 1;
Qy 48 CTGGGGATACCGGAGCACAACGGTCTTATTCACCTGGAAGGATTTTCCCTATTGCTCA 107
Db 273 CTGGGGGTACGGAAACAACCGGACTGAGCAGTGGCATAGGACTTCCCAATGCCAA 332
Qy 108 TGGTGATCAGCAATCTCCAAATGAGATTAAACCAAAGAAAGTGAATATGACTTTCCCT 167
Db 333 GGGAGAGCCGAGTCCCTGTTGACATCGACACTCATACAGCCAGTATGACCTTCCCT 392
Qy 168 CCGACCATTAGTATCAAGTATGACCCAGCTCAGCTAAATCATCAGCAACAGCGGCA 227
Db 393 GAAGCCCTGTCTGTTTCTATGATCAAGCAACTTCCCTGAGGATCCTCAACAATGGTCA 452
Qy 228 TTCTTCAATGTGATCTTTGATGACACAGAGACAAATCAGTTCTCGGTGGTGGTCTCT 287
Db 453 TGTCTTCAACGTGGAGTTTGATGACTCTCAGGACAAAGCAGTCTCAAGGAGGACCCCT 512
Qy 288 CACTGGAAGCTACAGGTTACCGCAGGTTACCTTTCATCTGGGGGTCCGCTGATGACCACGG 347
Db 513 GGATGGCACTTACAGATTGATTGATTTCACTTTCACTGGGGTTCATGATGACAAGG 572
Qy 348 CTCGAGCACAATAGTAGAGTGGATGATGCTGAGAGCTCCATGTTGTTCACTGGAA 407
Db 573 TTCAGAGCATCTGTGATGATAAAGAAATATGTCGAGAACTTCACTTGGTTCACTGGAA 632
Qy 408 TTCAGCAAAATACCCAGCTTTGTTGAGCAGCTCATGACCAAGATGAGCTGCTCTT 467
Db 633 CAC---CAATATATGGGATTTTGGGAAAGCTGTGACGAACCTGATGGACTGGCCGTTCT 689
Qy 468 GGGAGTGTGTTTACAGATTGGTGAACTTAATTCCTCAATTCGCAAAAGATTACTGACACTTT 527
Db 690 AGGTATTTTTTGAAGTTGGCAGCGCTAAACCCGGGCTTCAGAAAGTTGTTGATGTCT 749

Qy 528 GGATTCATTAAGAAAGGGTAAACAACTCGATTTCACAAATTTGACCTATTGCTCTCT 587
Db 750 GGATTCATTAAGAAAGGGCAGAGTGTGACTTCACTAACTTCGATCCTCGTGGCT 809
Qy 588 GCTTCCACCATCTCTGGGACTACTGGACATATCTTGGTCTTTACAGTTCCACCTCTTCT 647
Db 810 CTTCTCGAATCCCTGGATTACTGGACCTACCCAGGCTCACTGACCACCCCTCTCTTCT 869
Qy 648 TGAGAGTGTACATGGATTGTTTAAAGCAACCTATAAACAATCAGCTCTCAACAGCTGGC 707
Db 870 GGAATGTGTGACCTGGATTGTCTCAAGGAAACCCATCAGCGTCAGCAGCGAGCAGGTGT 929
Qy 708 CAAATTTGCGAGTCTCTGTCGACAGCGGAGGTTGAAGCAGCAG 751
Db 930 GAAATTCGGTAACTTAATTCATTTGGGAGGGTGAACCCGAAG 973

RESULT 10

US-09-938-270B-2
; Sequence 2, Application US/09938270B
; Patent No. 6673562
; GENERAL INFORMATION:
; APPLICANT: Qinwei Shi
; TITLE OF INVENTION: Differential Immunoassay
; FILE REFERENCE: 1112-1-080N
; CURRENT APPLICATION NUMBER: US/09/938,270B
; CURRENT FILING DATE: 2001-08-23
; PRIOR APPLICATION NUMBER: US 60/227,536
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/292,497
; PRIOR FILING DATE: 2001-05-21
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: homosapien
US-09-938-270B-2

Query Match 29.3%; Score 266.8; DB 4; Length 1266;
Best Local Similarity 62.0%; Pred. No. 5.5e-78;
Matches 440; Conservative 0; Mismatches 267; Indels 3; Gaps 1;
Qy 49 TGGGATACCGGAGCACAACGGTCTTATTCACCTGGAAGGATTTTCCCTATTGCTGAT 108
Db 496 TGGGGCTACGCGAGTCACAACGGTCTGACCACTGGCATGAACTTTTCCCAATGCCAAG 555
Qy 109 GGTGATCAGCAATCTCCAAATTCAGATTAAACCAAAGAAAGTGAATATGACTTCTCCCTC 168
Db 556 GGGGAAACCACTGCGCCCTTGAGCTGCATCTAAAGACATCAGGCATGACCCCTCTCTG 615
Qy 169 CGACCACTTATGATCAAGTATGACCCAGCTCAGCTAAATCATCAGCAACAGCGGCAAT 228
Db 616 CAGCCATGCTGTGTCTTATGATGTTGCTCTGCAAGACCATCTCTTAATATGGAAG 675
Qy 229 TCCTTCAATTTGACTTTTGGATGACACAGAGACAAATCAGTTCTGGGTGGTGGTCTCTC 288
Db 676 ACTGCGGAGTTGTTTATTTGATGATCTATGATAGTCAATGCTGAGAGGGGTCTCTC 735
Qy 289 ACTGGAAGCTACAGGTTACGGCAGGTTTCACTTTCCTGGGGGTCCGCTGATGACACCGG 348
Db 736 CTGGAACCTTACGACTTCGCCAGTTTCATCTTCACTGGGGCTCTTCGATGATCATGCG 795
Qy 349 TCCGAGCACAATAGTAGAGTGGATGAGCTATGCTGAGAGCTCCATGTTGTTCACTGGAAT 408
Db 796 TCTGAGCACAACCGTGGATGGAGTCAAGTATGAGCGGAGCTTCATTTGGTTCACTGGAAC 855
Qy 409 TCAGCAAAATACCCAGCTTTTGTGAGCAGCTCATGAACCAAGATGAGCTGGCTGTCTG 468
Db 856 CCG---AAGTATTAACACTTTTAAAGAAAGCCCTGGAAGCAGCGGATGGGATCGCTGTGAT 912
Qy 469 GGAGTCTGTTTACAGATTGGTGAACTTAATTCCTCAATTCGCAAAAGATTACTGACACTTTG 528

Db 913 GGCAATTTTCTGAAGATAGGACATGAGAAATGGCGAGTTCAGATTTTCTCTGATGATG 972
QY 529 GATTCATTAAGAAAGGTAACAACTCGATTACAAATTTTGAACCTATGTCCTG 588
Db 973 GACAGATTAAAGCAAGGCAAGGAGGCGCCTTCACAAAGTTTGAACCATCTCGCTG 1032
QY 589 CTTCACCACTCTCGGACTACTGGACATATCTCTGGTTCTCTTACAGTTCCACCTCTTCT 648
Db 1033 TTCCCGGATGCGGAGCTACTGGACTTACAGGCTCATTACACGCGCGCTCGGAG 1092
QY 649 GAGAGTGTCACTGATGATTTTAAAGCAACCTATATAACATCAGCTTCAACAGCTGGCC 708
Db 1093 GAATGCAATGTGTGGCTGCTCTCCAGTGTGAGAGGCCCATGACCGTAGCTCTGACCAAGTGGCC 1152
QY 709 AAATTCGAGTCTCTGTGTCACAGCGGAGGTGAAGCAGCAGCTTTTCT 758
Db 1153 AAGCTGGAGCGCTCTCTCCAGTGTGAGAGCGGCCCGCCAGTGCCTCT 1202

RESULT 11

US-09-385-982-446
; Sequence 446, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CDDNA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; CURRENT FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 446
; LENGTH: 599
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(599)
; OTHER INFORMATION: n = A, T, C or G
US-09-385-982-446

Query Match 19.5%; Score 177.4; DB 3; Length 599;
Best Local Similarity 61.0%; Pred. No. 1.5e-48;
Matches 353; Conservative 0; Mismatches 217; Indels 9; Gaps 4;
QY 53 GATACCGGAGCAACAGCTTCCTATTCTACCTGGAAGAAATTTTCCCTATTGCTGATGGTG 112
Db 1 GGTACGGCAACACAAACGACCTGAGCACTGGCATTAAGGACTTCCCAATTTGCCAAGGAG 60
QY 113 ATCAGCAATCTCCAATTGAGATTAAACCAAGAGTAAATATGACTTCTCCCTCCGAC 172
Db 61 AGCGCAGTCCCTGTGTGATCGACACTATACAGCCAAATGATGCCCTTCCCTGAAGC 120
QY 173 CACTTAGTATCAAGTATGACCCCAAGCTCAGCTAAAATCATCAGCAACAGCGGCCATTCCT 232
Db 121 CCTGTCTCTTTCTATGATCAAGCAACTTCCCTGAGGATCCTCAACAATGGTTCATGCTT 180
QY 233 TCATTTTCACTTTGATGACACAGAAACAAATCAGTTCTGCGTGGTGGTCTCTCTACTG 292
Db 181 TCAACGTGAGTTTGAATGACTCTCAGGACAAAGCAGTGTCAAGGGAGGACCCCTCGATG 240
QY 293 GAAGCTACAGGTTTACCGGAGGTTTCACTTCACTGGGGGTCCGCTGATGATCAGCGGCTCCG 352
Db 241 GCACCTACAGATTGATTCAGTTTCACTTTCACTGGGGTTCACCTTGTATGACAGGTTTCAT 300
QY 353 AGCACATAGTAGATGGAGTGAGCTATGCTGACAGAGCTCCATGTTGTTCACCTGGAATTCAG 412

Db 301 AGCATACTGTGATATAAAGAAATATGCTGCAGAACTTCACTTGGTTCTCACTGGAACAC-- 358
QY 413 ACAATATACCCAGCTTTTGTAGGAGCTCATGAACCAAGATGGAGCTGGTG-TCTTGGGA 471
Db 359 -CAAATATGGGATTTTGGGAAAGCTGTGAGCAACCTGATGAGCTGGCGGTTCTAGGTA 417
QY 472 GTGTTTTTACAGATTGGTGAACCTAATCCCAACTGCAAAAGATTACTGACACT-TTGGGA 530
Db 418 TTTTCTTTTGAAGGTTTGGCAGCGCTAAACNCGGCTTNAAGTTGTTGAATGCTGGA 477
QY 531 TTCCATTAAGAAAGGGTAAACAACTCGATTACAAAATTTTGACCTATTGCTCTGCT 590
Db 478 TTCCATTAAGAAAGGGCAAGAAATTGCTGACTTTTCACTAATTNNAAATCTCTGNGGCT 537
QY 591 TCCACC-----ATCCTGGGACTACTGACATATCTCGTT 625
Db 538 TCTTCTGAAATCCTTTGGATTACCGGACCTNCCAGCTT 576

RESULT 12

US-09-621-976-18059
; Sequence 18059, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 18059
; LENGTH: 477
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-18059

Query Match 16.2%; Score 147.6; DB 4; Length 477;
Best Local Similarity 65.6%; Pred. No. 9.4e-39;
Matches 244; Conservative 1; Mismatches 125; Indels 2; Gaps 2;
QY 51 GGGATACCGGAGCAACAGCTCCTATTCTACCTGGAAGGAATTTTCCCTATTGCTCATGG 110
Db 66 GGGGCTACCGAGTCACACGGTCTGACCACTGGCATGAACCTTTTCCCAATGCCAAGG 125
QY 111 TGATCAGCAATCTCCAATTGAGATTAAACCAAGAGTGAATATGACTTCTTCCCTCCG 170
Db 126 GGAACACCAAGTCGCCCTTTGAGCTGCATACTAAAGACATCAGCATGACCTTCTCTGCA 185
QY 171 ACACCTTAGTATCAAGTATGACCCAGCTCAGTAAATCATCAGCAACAGCGGCATTC 230
Db 186 GCCATGGTGTGTCTTATGATGGTGGCTCTGC-AAGACCATCTCTGAATAATGGGAAGAC 244
QY 231 CTTCAATGTTGACTTTTGTATGACACAGAGAAACAAATCAGTTCTGCGTGGTGGTCTCTCAC 290
Db 245 CTGCGAGTTGTTATTTGATGATCTTATGATAGTCAATGCTGAGAGGGGTTCTCTCCC 304
QY 291 TGAAGCTACAGGTTTACCGCAGGTTTCACTTCACTGGGGGTCCGCTGATGACCAACGCTC 350
Db 305 TGAACCTTACCGACTTCGCGAGTTTTCATCTTCACTGGGGCTCTTCGGATGATCATGGCTC 364
QY 351 CGAGCATAGTAGATGGAGTGAAGCTATGCTGAGAGCTTCCATGTTGTTCACTGGAATTC 410
Db 365 TGAGCACACCGTGGATGGAGTCAAGTATCAGCGGAG-TTCATTTGTTGTTCACTGGAACCC 423
QY 411 AGACAATAACCC 422
Db 424 GAAGTATAACAC 435

```
RESULT 13
US-09-385-982-136
; Sequence 136, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCDNA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; CURRENT FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 136
; LENGTH: 581
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(581)
; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-136

Query Match
Best Local Similarity 16.2%; Score 147.4; DB 3; Length 581;
Matches 214; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

QY 419 ACCCAGCTTTGTGTGAGGAGCTCATGAACACAGATGGCTGCTGCTTGGGAGTGTCTT 478
Db 1 ACTCCAGCTTGTGAGGCTGCTCAAGCTGCTCAAGGCTGATGTTTGGCAGTATTGGTGTGCA 60
QY 479 TACAGATTGGTGAACCTTAATTCCTCAACTGCAAAAGATTACTGACACTTTGGATTCCATTA 538
Db 61 TGAAGGTTGGTGGAGGCAACCCCAAGCTGCAGAAAGTACTTGATGCCCTCCCAAGCAATTA 120
QY 539 AGAAGAGGTAACAACTCGATTACAAATTTTACCAATTTTACCTATGTTCTCTGCTCCACCAT 598
Db 121 AAACCAAGGCAACAGAGCCCAATCACAATTTTACCCCTCTACTCTCTTCTCTTCAT 180
QY 599 CCTGGACTACTGACATATCTCTGTTCTTTACAGTTCCACCTCTTTTGAGAGTGTCA 658
Db 181 CCTGGATTCTGGACTACCTGGCTCTCTGACTCATCTCTCTTTATGAGAGTGA 240
QY 659 CATGATTGTTTAAAGCAACCTATAAATCATGCTCTCAAGCTGGCCAAATTTCCGA 718
Db 241 CTTGGATCATCTGTAAGGAGAGCATCAGTGTCACTCAGAGCAGCTGGCACAATTCGCA 300
QY 719 GTCTCTGTGCACAGCGGAGGTGA 743
Db 301 GCCTTCTCAATGTTGAAGGTGA 325

RESULT 14
US-09-385-982-156/c
; Sequence 156, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCDNA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; CURRENT FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27

; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 156
; LENGTH: 760
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(760)
; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-156

Query Match
Best Local Similarity 15.0%; Score 136.8; DB 3; Length 760;
Matches 198; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 30 GACCATGTCGAGGCTCAGCTGGGATACCGCAGACACACGCTCTCTTCTTCACTTGAAGCA 89
Db 310 GATAATGGCAAGTCCAGACTGGGATATGATGACAAAATGGTCTCTGAAACAATGGAGCA 251
QY 90 ATTTTCCCTATTGCTGATGTCATCAGCAATCTCCAATTTGAGATTAAACCAAGAACT 149
Db 250 GCTGTATCCATTGCCAATGGAAATAACCAAGTCCCTCTGTTGATATTAAACCAAGTGA 191
QY 150 GAAATATGACTCTTCCCTCCGACCACTTAGTATCAAGTATGACCCAGCTCAGCTTAAAT 209
Db 190 CAAACATGACACCTCTCTGAAACCTATTAGTGTCTCTACAACCCAGCCACCAAGA 131
QY 210 CATCAGCACACGCGCCATTCTTCAATGTTTCACTTTGATGACACAGAACAAATCACT 269
Db 130 AATTATCAATGTTGGGCACTTCTTCATGTAATTTTGGAGCAACGATACCGATCAGT 71
QY 270 TCTGCTGGTGTGCTCTCTCACTGGAAGCTACAGGTTACGCGAGGTTCACTTCACTGGG 329
Db 70 GCTGAAAGTGTGCTCTTCTCTGACAGCTACAGGCTCTTTTCACTTCCATTTTCACTGGG 11

RESULT 15
US-09-385-982-472
; Sequence 472, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCDNA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; CURRENT FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 472
; LENGTH: 586
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(586)
; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-472

Query Match
Best Local Similarity 12.2%; Score 111; DB 3; Length 586;
Matches 174; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

QY 519 TGACACTTTGGATTTCATTAAAGAAAGGTAAACAACTCGATTCAAAATTTTGACCT 578
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 2004, 22:47:51 ; Search time 499 Seconds
(without alignments)
9185.806 Million cell updates/sec

Title: US-10-069-434-4
Perfect score: 911
Sequence: 1 cggattcgctcgagttcc.....cccaatgccagggggaaaa 911

Scoring table: IDENTITY_NUC
Gapop 10_0 , Gapext 1.0

Searched: 3304383 seqs, 2515761380 residues

Total number of hits satisfying chosen parameters: 6608766

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA.*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
- 17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	911	100.0	911	15	US-10-069-434-4
2	759	83.3	1248	13	Sequence 4, Appli
3	745.8	81.9	1023	13	Sequence 4, Appli
4	745.8	81.9	3564	16	Sequence 19, Appl
5	444.4	48.8	496	10	Sequence 639, App
6	331.2	36.4	2785	9	Sequence 14076, A
7	331.2	36.4	2785	15	Sequence 3, Appli
8	331.2	36.4	2785	15	Sequence 31, Appl
9	329.6	36.2	1233	15	Sequence 31, Appl
10	329.6	36.2	1233	15	Sequence 360, App
11	329.6	36.2	1244	9	Sequence 79, Appl
12	329.6	36.2	1244	15	Sequence 23, Appl
13	289.4	31.8	1459	9	Sequence 47, Appl
14	289.4	31.8	1459	16	Sequence 1675, Ap
					Sequence 124, App

15	284.8	31.3	605	13	US-09-988-292-17	Sequence 17, Appl
16	284.8	31.3	605	17	US-10-776-601-17	Sequence 17, Appl
17	280	30.7	1759	9	US-09-981-353-116	Sequence 116, App
18	280	30.7	1759	13	US-10-116-802-234	Sequence 234, App
19	278.4	30.6	1523	9	US-09-954-456-287	Sequence 287, App
20	278.4	30.6	1523	9	US-09-880-107-3844	Sequence 3844, Ap
21	278.4	30.6	1523	10	US-09-873-367C-825	Sequence 825, App
22	278.4	30.6	1523	12	US-09-968-007A-473	Sequence 473, App
23	278.4	30.6	1551	13	US-10-342-887-423	Sequence 423, App
24	278.4	30.6	1551	13	US-10-172-118-423	Sequence 423, App
25	278.4	30.6	1551	15	US-10-102-524-1753	Sequence 1753, App
26	278.4	30.6	1723	9	US-09-925-299-172	Sequence 172, App
27	278.4	30.6	1723	10	US-09-925-299-172	Sequence 172, App
28	278.4	30.6	1755	15	US-10-158-646-46	Sequence 46, Appl
29	278.4	30.6	1755	15	US-10-240-965-199	Sequence 199, App
30	278.4	30.6	1886	10	US-09-814-353-21737	Sequence 21737, A
31	278.4	30.6	2391	10	US-09-822-846-103	Sequence 103, App
32	266.8	29.3	1266	9	US-09-938-270B-2	Sequence 2, Appli
33	266.8	29.3	2286	13	US-10-363-616-5	Sequence 5, Appli
34	259.8	28.5	1053	16	US-10-388-934-93	Sequence 93, Appli
35	255	28.0	988	9	US-09-917-800A-1676	Sequence 1676, Ap
36	247	27.1	247	17	US-10-363-829-46	Sequence 46, Appl
37	223	24.5	1508	15	US-10-084-817-11	Sequence 11, Appl
38	198.8	21.8	1825	13	US-10-302-172-619	Sequence 619, App
39	197.8	21.7	507	9	US-09-981-353-102	Sequence 102, App
40	197.6	21.7	1201	9	US-09-917-800A-1677	Sequence 1677, Ap
41	197.6	21.7	1201	16	US-10-388-934-507	Sequence 507, App
42	184.8	20.3	421	9	US-09-960-352-1323	Sequence 1323, App
43	177.4	19.5	599	10	US-09-871-161-446	Sequence 446, App
44	174.4	19.1	683	17	US-10-404-460-38	Sequence 38, Appl
45	166.2	18.2	428	9	US-09-960-352-566	Sequence 566, App

ALIGNMENTS

RESULT 1

US-10-069-434-4
; Sequence 4, Application US/10069434
; Publication No. US20030121061A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: THORNTON, Michael
; APPLICANT: RAMKOMAR, Jayalaxmi
; APPLICANT: TRIBOULEY, Catherine M.
; APPLICANT: YUE, Henry
; APPLICANT: NGUYEN, Damiel B.
; APPLICANT: YAO, Monique G.
; APPLICANT: PATTERSON, Chandra
; APPLICANT: GANDHI, Ameena R.
; APPLICANT: BURFORD, Neil
; APPLICANT: THANGAVELU, Kavitha
; APPLICANT: BAUGHN Mariah R.
; TITLE OF INVENTION: HUMAN LYASES
; FILE REFERENCE: PI-0137 PCT
; CURRENT APPLICATION NUMBER: US/10/069,434
; CURRENT FILING DATE: 2002-02-20
; PRIOR APPLICATION NUMBER: 60/213,383; 60/215,544; 60/222,818
; PRIOR FILING DATE: 2000-06-23; 2000-06-30; 2000-08-04
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PERL Program
; SEQ ID NO 4
; LENGTH: 911
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20030121061A1 6338333CB1
US-10-069-434-4

Query Match 100.0%; Score 911; DB 15; Length 911;
Best Local Similarity 100.0%; Pred. No. 1.3e-283;
Matches 911; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 CGGAATTCGGCTCGAGTTCCACCCGAGGACCATGTGAGGCTCAGCTGGGATACCGC 60
Db 1 CGGAATTCGGCTCGAGTTCCACCCGAGGACCATGTGAGGCTCAGCTGGGATACCGC 60
QY 61 GAGCACAACGGTCTCTATTCACTGGAAGAAATTTTCCCTATTGCTGATGTGATCAGCAA 120
Db 61 GAGCACAACGGTCTCTATTCACTGGAAGAAATTTTCCCTATTGCTGATGTGATCAGCAA 120
QY 121 TCTCCAATTGAGATTAAAAACAAAGAGTGAATATGACTCTTCCCTCCGACCATTAGT 180
Db 121 TCTCCAATTGAGATTAAAAACAAAGAGTGAATATGACTCTTCCCTCCGACCATTAGT 180
QY 181 ATCAAGTATGACCCAGCTCAGCTTAAATCATCAGCAACAGCGGCCATTCCTTCAATGTT 240
Db 181 ATCAAGTATGACCCAGCTCAGCTTAAATCATCAGCAACAGCGGCCATTCCTTCAATGTT 240
QY 241 GACTTTGATGACACAGAGAAACAATCAGTTCTGCGTGGTGGTCTCTCACTGGAAGCTAC 300
Db 241 GACTTTGATGACACAGAGAAACAATCAGTTCTGCGTGGTGGTCTCTCACTGGAAGCTAC 300
QY 301 AGTTAGCGGAGGTTTCACTTCACTGGGGGTCCGCTGATGACCAACGGCTCCGAGCACATA 360
Db 301 AGTTAGCGGAGGTTTCACTTCACTGGGGGTCCGCTGATGACCAACGGCTCCGAGCACATA 360
QY 361 GTAGATGGAGTGAGTATGCTGAGAGCTCCATGTTGTTCACTGGAATTCAGACAAATAC 420
Db 361 GTAGATGGAGTGAGTATGCTGAGAGCTCCATGTTGTTCACTGGAATTCAGACAAATAC 420
QY 421 CCCAGCTTTGTTGAGGAGCTCATGAACAGATGGAGTGGCTGTCTTGGAGTGTGTTTAA 480
Db 421 CCCAGCTTTGTTGAGGAGCTCATGAACAGATGGAGTGGCTGTCTTGGAGTGTGTTTAA 480
QY 481 CAGATTGGTGAACTTAATCCCACTGCAAAAGATTACTGACACTTTGGATTCCATTAAA 540
Db 481 CAGATTGGTGAACTTAATCCCACTGCAAAAGATTACTGACACTTTGGATTCCATTAAA 540
QY 541 GAAAAGGGTAAACAAACTCGANTTACAAAATTTTGACCTATTGCTCTGCCATCCATCC 600
Db 541 GAAAAGGGTAAACAAACTCGANTTACAAAATTTTGACCTATTGCTCTGCCATCCATCC 600
QY 601 TGGGACTACTGGACATATCTGTTCTTCTACAGTTCCACCTCTCTTGAGAGTGTCA 660
Db 601 TGGGACTACTGGACATATCTGTTCTTCTACAGTTCCACCTCTCTTGAGAGTGTCA 660
QY 661 TGGAATGTTTTAAAGCAACCTATAACATCAGCTCTCAACAGCTGGCCAAATTTGGAGT 720
Db 661 TGGAATGTTTTAAAGCAACCTATAACATCAGCTCTCAACAGCTGGCCAAATTTGGAGT 720
QY 721 CTCCTGTGACAGCGGAGGTGAAGCAGCAGCTTTTCTGTGATAGAGTCTCACTCTGTCA 780
Db 721 CTCCTGTGACAGCGGAGGTGAAGCAGCAGCTTTTCTGTGATAGAGTCTCACTCTGTCA 780
QY 781 CCCAGGCTGGAGGCGAGTGTGATCAATCTTGGCTTAATTGACGCTCCAACTCTCTGGACTCA 840
Db 781 CCCAGGCTGGAGGCGAGTGTGATCAATCTTGGCTTAATTGACGCTCCAACTCTCTGGACTCA 840
QY 841 AGTGATCTCCCACTCAGCTCAGAGTCTGACAGTCTGACCACTGGGATGATCTTTCCCAATGC 900
Db 841 AGTGATCTCCCACTCAGCTCAGAGTCTGACCACTGGGATGATCTTTCCCAATGC 900
QY 901 CAGGGGGGAAAA 911
Db 901 CAGGGGGGAAAA 911
```

RESULT 2

US-10-220-120-4

; Sequence 4, Application US/10220120
; Publication No. US20040048253A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: PANZER, Scott R.

```
; APPLICANT: SPIRO, Peter A.  
; APPLICANT: BANVILLE, Steven C.  
; APPLICANT: SHAH, Purvi  
; APPLICANT: CHALUP, Michael S.  
; APPLICANT: CHANG, Simon C.  
; APPLICANT: CHEN, Alice  
; APPLICANT: D'SA, Steven A.  
; APPLICANT: AMSHEY, Stefan  
; APPLICANT: DAHL, Christopher R.  
; APPLICANT: DAM, Tam C.  
; APPLICANT: DANIELS, Susan E.  
; APPLICANT: DUFOUR, Gerard E.  
; APPLICANT: FLORES, Vincent  
; APPLICANT: FONG, Willy T.  
; APPLICANT: GREENAWALT, Lila B.  
; APPLICANT: HILLMAN, Jennifer L.  
; APPLICANT: JONES, Anissa L.  
; APPLICANT: LIU, Tommy P.  
; APPLICANT: ROSEBERRY, Ann M.  
; APPLICANT: ROSEN, Bruce H.  
; APPLICANT: RUSSO, Frank D.  
; APPLICANT: STOCKREHER, Theresa K.  
; APPLICANT: DAFFO, Abel  
; APPLICANT: WRIGHT, Rachel J.  
; APPLICANT: YAP, Pierre E.  
; APPLICANT: YU, Jimmy Y.  
; APPLICANT: BRADLEY, Diana L.  
; APPLICANT: BRATCHER, Shawn R.  
; APPLICANT: CHEN, Wensheng  
; APPLICANT: COHEN, Howard J.  
; APPLICANT: HODGSON, David M.  
; APPLICANT: LINCOLN, Stephen E.  
; APPLICANT: JACKSON, Stuart  
; TITLE OF INVENTION: MOLECULES FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: PT-1113 PCT  
; CURRENT APPLICATION NUMBER: US/10/220,120  
; CURRENT FILING DATE: 2002-08-26  
; PRIOR APPLICATION NUMBER: 60/184,777; 60/184,797; 60/184,698; 60/184,770; 60/184,774,  
60/184,693; 60/184,771; 60/184,813; 60/184,773; 60/184,776;  
60/184,769; 60/184,768; 60/184,837; 60/184,697; 60/184,841;  
60/184,772; 60/185,213; 60/185,216; 60/204,863; 60/205,221;  
60/204,815; 60/203,785; 60/204,821; 60/204,908; 60/204,226;  
60/204,525; 60/205,285; 60/205,232; 60/205,323; 60/205,287;  
60/205,324; 60/205,286  
; PRIOR FILING DATE: 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24;  
2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24;  
2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24;  
2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24;  
2000-05-17; 2000-05-12; 2000-05-16; 2000-05-16; 2000-05-15;  
2000-05-16; 2000-05-17; 2000-05-16; 2000-05-17; 2000-05-17;  
2000-05-17; 2000-05-17  
; NUMBER OF SEQ ID NOS: 422  
; SOFTWARE: PERL Program  
; SEQ ID NO 4  
; LENGTH: 1248  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. US20040048253A1 LI:090574.1:2000FEB01  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: 1148, 1234  
; OTHER INFORMATION: a, t, c, g, or other  
US-10-220-120-4
```

Query Match 83.3%; Score 759; DB 13; Length 1248;

Best Local Similarity 100.0%; Pred. No. 2.5e-234;

Matches 759; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGAATTCGGCTCGAGTTCCACCCGAGGACCATGTGAGGCTCAGCTGGGATACCGC 60

Db 15 CGGAATTCGGCTCGAGTTCCACCCGAGGACCATGTGAGGCTCAGCTGGGATACCGC 74


```
|||||
Db 411 CAGAGCTCAATGTTGTTCACTGGAATTCAGACAAATACCCAGCTTTGTTAGGCGATC 470
QY 443 ATGAACCAAGATGGACTGGCTGTCCTTG 468
Db 471 ATGAACCAAGATGGACTGGCTGTCCTTG 496

RESULT 6
US-09-802-674-3
; Sequence 3, Application US/09802674
; Patent No. US20020042088A1
; GENERAL INFORMATION:
; APPLICANT: Macina, Roberto A
; APPLICANT: Piderit, Alejandra
; APPLICANT: Sun, Yongming
; TITLE OF INVENTION: Method of Diagnosing, Monitoring, Staging, Imaging and
; TITLE OF INVENTION: Treating Gastrointestinal Cancer
; FILE REFERENCE: DEX-0142
; CURRENT APPLICATION NUMBER: US/09/802,674
; CURRENT FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/188,061
; PRIOR FILING DATE: 2000-03-09
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 2785
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-802-674-3

Query Match 36.4%; Score 331.2; DB 9; Length 2785;
Best Local Similarity 64.5%; Pred. No. 1.2e-95;
Matches 495; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

QY 30 GACCATGTCGAGGCTCAGCTGGGGATACCGCGAGCACAACGGTCTCTATTCACTGGGAAGCA 89
Db 1044 GATATGGCAAGTCAGACTGGGGATATGATGACAAATAATGCTCTGGAACAATGGAGCAA 1103
QY 90 ATTTTTCCTATTGCTGATGATCAGCAATCTCCAATTGAGATTAAACCAAAAGAACT 149
Db 1104 GCTGTATCCATTGCCAATGGAATAACCAATCCCTCTGTTGATATTAAACCCAGTGAAC 1163
QY 150 GAAATATGACTCTTCCCTCCGACCACTTAGTATCAAGTATGACCAAGCTCAGCTAAAT 209
Db 1164 CAACATGACACTCTCTGAAACCTTAGTGTCTCTACACCCAGCCACGACCCANAGA 1223
QY 210 CATCAGCAACAGCGGCCATTCTTCAATGTTGACTTTGATGACACAGAGAACAAATCACT 269
Db 1224 GCTGAAAGGTGGTCTCTCTGACAGCTACAGGCTCTTTCAGTTTCAATTTTCACTGGGG 1343
QY 330 GTCGGTGTGTCCTCTCACTGGAGCTACAGGTTACGGCAGGTTCACTTCACTGGGG 329
Db 1344 CAGTACAAATGAGCATGGTTCAGAACATACAGTGGATGGAGTCAATATTCTGCCGAGCT 1403
QY 390 CCATGTTGTTCACTGGAATTCAGACAAATACCCAGCTTTTACAGATTGGTGAACCTTAATCCCACTGCA 449
Db 1404 TCACGTAGCTCACTGGAATTCGCAAAAGTCTCCAGCTTCTGCTGAAGCTGCTCTCAAAGGC 1463
QY 450 AGATGAGCTGGCTGCTCTGGGAGTGTTTTACAGATTGTTGAGGAGCTGATGAACCTGCA 509
Db 1464 TGAATGTTGGCAGTTATTGGTGTGTTGATGAAGGTTGGTGGGCCCAACCCAAAGCTGCA 1523
QY 510 AAGATTACTGACACTTTGGAATTCATTAAGAAAAGGGTAAACAACTCGATTTCACAA 569
Db 1524 GAAAGTACTTGATGGCTCTCAAGCAATTTAAACCAAGGGCAACGAGGCCCACTTCAAAA 1583
QY 570 TTTTGAACCTATTGTTCTGCTCCACCTCCTGGGACTACTGGACATATCCTGGTCTCT 629
|||||

Db 1584 TTTTGACCCCTTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1643
QY 630 TACAGTTTCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 689
Db 1644 GACTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1703
QY 690 CAGCTCTCAACAGCTGGCCAAATTTTCGCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 749
Db 1704 CAGCTCAGAGCAGCTGGCACAATTCGCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1763
QY 750 AGCTTTTCTGTGATAGAGTCTCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 797
Db 1764 TGTCCCATGTCAGCACAACAACCGCCCAACCACTCTCTGAAGGCGAG 1811

RESULT 7
US-10-393-892-31
; Sequence 31, Application US/10393892
; Publication No. US20030186302A1
; GENERAL INFORMATION:
; APPLICANT: WANG, YIXIN
; TITLE OF INVENTION: COLORECTAL CANCER DIAGNOSTICS
; FILE REFERENCE: CDS 267 US NP
; CURRENT APPLICATION NUMBER: US/10/393,892
; CURRENT FILING DATE: 2003-03-21
; PRIOR APPLICATION NUMBER: 60/368,798
; PRIOR FILING DATE: 2002-03-29
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 31
; LENGTH: 2785
; TYPE: DNA
; ORGANISM: human
US-10-393-892-31

Query Match 36.4%; Score 331.2; DB 15; Length 2785;
Best Local Similarity 64.5%; Pred. No. 1.2e-95;
Matches 495; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

QY 30 GACCATGTCGAGGCTCAGCTGGGGATACCGCGAGCACAACGGTCTCTCTATTCACTGGGAAGCA 89
Db 1044 GATAATGGCAAGTCCAGACTGGGGATATGATGACAAATAATGCTCTCTGAAACAATGGAGCAA 1103
QY 90 ATTTTTCCTATTGCTGATGATCAGCAATCTCCAATTGAGATTAAACCAAAAGAACT 149
Db 1104 GCTGTATCCATTGCCAATGGAATAACCAATCCCTCTGTTGATATTAAACCCAGTGAAC 1163
QY 150 GAAATATGACTCTTCCCTCCGACCACTTAGTATCAAGTATGACCAAGCTCAGCTAAAT 209
Db 1164 CAACATGACACTCTCTGAAACCTTAGTGTCTCTACACCCAGCCACGACCCANAGA 1223
QY 210 CATCAGCAACAGCGGCCATTCTTCAATGTTGACTTTGATGACACAGAGAACAAATCACT 269
Db 1224 AATTATCAATGTTGGGCACTTCTTCCATGTAATTTTGAGGACAACGATAACCGATCACT 1283
QY 270 TCTGGTGTGTCCTCTCACTGGAGCTACAGGTTACGGCAGGTTCACTTCACTGGGG 329
Db 1284 GCTGAAAGGTGGTCTCTCTCTGACAGCTACAGGCTCTTTCAGATTTCATTTTCACTGGGG 1343
QY 330 GTCGGTGTGATGACCGGCTCCGAGCACAATAGTAGTGGAGTGAGCTATGCTGCGAGAGCT 389
Db 1344 CAGTACAAATGAGCATGGTTCAGAACATACAGTGGATGGAGTCAATATTCTGCCGAGCT 1403
QY 390 CCATGTTGTTCACTGGAATTCAGACAAATACCCAGCTTTTACAGATTGGTGAACCTTAATCCCACTGCA 449
Db 1404 TCACGTAGCTCACTGGAATTCGCAAAAGTCTCCAGCTTCTGCTGAAGCTGCTCTCAAAGGC 1463
QY 450 AGATGAGCTGGCTGCTCTGGGAGTGTTTTACAGATTGTTGAGGAGCTGATGAACCTGCA 509
Db 1464 TGAATGTTGGCAGTTATTGGTGTGTTGATGAAGGTTGGTGGGCCCAACCCAAAGCTGCA 1523
QY 510 AAGATTACTGACACTTTGGAATTCATTAAGAAAAGGGTAAACAACTCGATTTCACAA 569
Db 1524 GAAAGTACTTGATGGCTCTCAAGCAATTTAAACCAAGGGCAACGAGGCCCACTTCAAAA 1583
QY 570 TTTTGAACCTATTGTTCTGCTCCACCTCCTGGGACTACTGGACATATCCTGGTCTCT 629
|||||
```

Db 1524 GAAAGTACTGTATGTCCTCCAAAGCAATTAAACCAAGGCAACAGAGCCCACTTACARA 1583
QY 570 TTTTGGACCTATTGTCTCTGCTTCCACCATCTCTGGGACTACTGGACATATCTCTGTTCTCT 629
Db 1584 TTTTGGACCCCTCTACT 1643
QY 630 TACAGTTCCACCTCTCTCTGAGAGTGTCAATGGATTGTTTAAAGCAACCTTATAACAT 689
Db 1644 GACTCATCTCTCTCTTTATGAGAGTGTAACTTGGATCATCTGTAAGGAGAGCATCAGTGT 1703
QY 690 CAGCTCTCAACAGCTGGCCAAATTTCCAGTCTCTCTGTGCAAGCGGAGGGTGAAGCAGC 749
Db 1704 CAGCTCAGAGCAGCTGGCAATTTCCGAGCCTTCTATCAAAATGTTGAAGGTGATAACGC 1763
QY 750 AGCTTTTCTGTGATAGTCTCACTCTGTCAACCCAGGCTGGAGGGCAG 797
Db 1764 TGTCCCCATGCAACAAACCGCCCAACCTCTGAAGGGCAG 1811

RESULT 8

US-10-394-382-31
; Sequence 31, Application US/10394382
; Publication No. US20030186303A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Yixin
; TITLE OF INVENTION: COLORECTAL CANCER DIAGNOSTICS
; FILE REFERENCE: CDS 266 US NP
; CURRENT APPLICATION NUMBER: US/10/394,382
; PRIOR FILING DATE: 2003-03-21
; PRIOR APPLICATION NUMBER: 60/368,687
; PRIOR FILING DATE: 2002-03-29
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 31
; LENGTH: 2785
; TYPE: DNA
; ORGANISM: human
US-10-394-382-31

Query Match 36.4%; Score 331.2; DB 15; Length 2785;
Best Local Similarity 64.5%; Pred. No. 1.2e-95;
Matches 495; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

QY 30 GACCATGTCCAGCTCAGCTGGGGATACCGGAGCAACAGGTCCTATTCACTGGAGGA 89
Db 1044 GATAATGGCAAGTCCAGACTGGGGATATGATGACAAAATGGTCTGAAATGGAGCAA 1103
QY 90 ATTTTTCCTATTGTCTGATGGTGTATCAGCAATCTCCAAATGAGATTAAACCAAGAAAGT 149
Db 1104 GCTGTATCCATTTGCCAATGGAAATAACCAATCCCTCTGTTGATTTAAACCAAGTGAAC 1163
QY 150 GAAATATGACTTTCCTCGGACCATCTAGTATCAAGATGACCCAGCTCAGCTAAAT 209
Db 1164 CAAACATGACACCTCTCTGAAACCTATTAGTGTCTCTACAAACCCAGCCACAGCCAAAGA 1223
QY 210 CATCAGCAACAGCGGCCATTTCCTCAATGTTGACTTTTGATGACACAGAGAAACATCAGT 269
Db 1224 AATATCAATGTGGGCAATCTTCCATGTAATTTTGGAGCAACGATACCGATCAGT 1283
QY 270 TCTGCGTGTGTCTCTCTCACTGGAAGCTACAGGTTACGGCAGGTTCACTTTCCTGGGG 329
Db 1284 GCTGAAAGGTGGTCTTCTCTGACAGCTACAGGCTCTTTCAGTTTCATTTTTCCTGGGG 1343
QY 330 GTCCGCTGATGACACAGGCTCGAGCAGCATAGTAGAGTGGAGTGTGCTGAGAGCT 389
Db 1344 CAGTACAAATGAGCATGGTTTCAGAACATACAGTGGATGGAGTCAAAATATTCTGCCGAGCT 1403
QY 390 CCATGTTGTTTCACTGGAATTCAGACAAATACCCAGCTTTGTTGAGCGACTCATGAACC 449
Db 1404 TCAGTAGCTCACTGGAAATTCGAAAGTACTCCAGGCTTGTGAGAGTGCCTCAAGGC 1463
QY 450 AGATGAGCTGGCTGTCTTGGAGTGTGTTTACAGATTGGTGAACCTTAATTCCTCAACTGGA 509

Db 1464 TGATGTTTTGGCAGTTATTGGTGTGTTGATGAAGGTTGGTGAGGCCAACCCAAAGCTGCA 1523
QY 510 AAAGATTACTGACACTTTGGATTCCATTTAAAGAAAGGGTAAACAACTCGATTTCAAAA 569
Db 1524 GAAAGTACTTGTATGTCCTCCAAAGCAATTAAACCAAGGCAACAGAGCCCAATTCACAA 1583
QY 570 TTTTGGACCTATTGTCTCTGCTTCCACCATCTCTGGGACTACTGGACATATCTCTGGTTCTCT 629
Db 1584 TTTTGGACCCCTCTACT 1643
QY 630 TACAGTTCCACCTCTCTTTGAGAGTGTCAATGGATTGTTTAAAGCAACCTTATAACAT 689
Db 1644 GACTCATCTCTCTCTTTATGAGAGTGTAACTTGGATCATCTGTAAGGAGAGCATCAGTGT 1703
QY 690 CAGCTCTCAACAGCTGGCCAAATTTCCAGTCTCTCTGTGCAAGCGGAGGGTGAAGCAGC 749
Db 1704 CAGCTCAGAGCAGCTGGCAATTTCCGAGCCTTCTATCAAAATGTTGAAGGTGATAACGC 1763
QY 750 AGCTTTTCTGTGATAGTCTCACTCTGTCAACCCAGGCTGGAGGGCAG 797
Db 1764 TGTCCCCATGCAACAAACCGCCCAACCTCTGAAGGGCAG 1811

RESULT 9

US-10-106-698-360
; Sequence 360, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 360
; LENGTH: 1233
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1222)..(1222)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-106-698-360

Query Match 36.2%; Score 329.6; DB 15; Length 1233;
Best Local Similarity 64.3%; Pred. No. 2.4e-95;
Matches 494; Conservative 0; Mismatches 274; Indels 0; Gaps 0;

QY 30 GACCATGTCCAGCTCAGCTGGGGATACCGGAGCAACAGGTCCTATTCACTGGAGGA 89
Db 73 GATAATGGCAAGTCCAGACTGGGGATATGATGACAAAATGGTCTGAAATGGAGCAA 132
QY 90 ATTTTTCCTATTGTCTGATGGTGTATCAGCAATCTCCAAATGAGATTAAACCAAGAAAGT 149
Db 133 GCTGTATCCATTTGCCAATGGAAATAACCAAGTCCCTCTGATATATTAACCAAGTGAAC 192
QY 150 GAAATATGACTCTTCCCTCGGACCATCTAGTATCAAGTATGACCCAGCTCAGCTAAAT 209
Db 193 CAAACATGACACCTCTCTGAAACCTATTAGTGTCTCTACAAACCCAGCCACAGCCAAAGA 252
QY 210 CATCAGCAACAGCGGCCATTTCCTTCAATGTTGACTTTTGATGACACAGAGAAACATCAGT 269
Db 253 AATATCAATGTGGGCAATTCCTTCCATGTAATTTTGGAGCAACGATACCGATCAGT 312
QY 270 TCTGCGTGTGTGTCCTCTCACTGGAAGCTACAGGTTACGGCAGGTTCACTTTCACCTGGGG 329

313	Db	GCTGAAAGGTGGTCCTTTCTCTGACAGCTACAGGCTCTTTTCAGTTTCCATTTTTCACCTGGG	372
330	Qy	GTCGGCTGATGACCAACGGCTCCGAGCACATAGTAGATGGAGTGAGCTATGCTGCAGAGCT	389
373	Db	CAGTACAAATGAGCATGGTTTCAGAAACATACAGTGGATGGAGTCAAAATATCTGCCGAGCT	432
390	Qy	CCATGTGTTTCACGTGGAAATCAGACAAATACCCAGCTTTCTTGGGCGAGCTCATGAACC	449
433	Db	TCACGTAGCTCACTGGAAATCTCGAAAGTACTCCAGCCCTTCTGAAGCTGGCTCAAAGGC	492
450	Qy	AGATGGACTGGCTGTCTTTGGGAGTGTTTTTACAGATTGGTGAACCTAATTCGCCAACTGCA	509
493	Db	TGATGTTTGGCAGTTATTGCTGTTTGGATGAAGCTTGGTCAGGCCAACCCAAAGCTGCA	552
510	Qy	AAAGATTACTGACACTTTGGATTCATTAAGAAAGGGTAAACAAACTCGATTACAAAA	569
553	Db	GAAAGTACTTTGATGCCCTCCAAGCAATTTAAACCAAGGGCAACGAGGCCCATTCACAAA	612
570	Qy	TTTTGACCTATGTCTCTGCTTCCACCATCCTGGGACTACTGGACATATCTCTGGTTCTCT	629
613	Db	TTTTGACCCCTCTACTCTCCTTCTCTCATCCCTGGATTTCTGGACCTTACCCTGGCTCTCT	672
630	Qy	TACAGTTCCACCTCTCTTGAGAGTGTCACATGGANTGTTTTAAAGCAACCTATAAACAAT	689
673	Db	GACTCATCTCCTCTTTATGAGAGTGTAACTTGGATCATCTGTGAAGGAGAGCATCAGTGT	732
690	Qy	CAGCTCTCAAACAGCTGGCCAAATTTCCGAGTCTCTCTGTGCACAGCGGAGGGTGAAGCAGC	749
733	Db	CAGCTCAGAGCAGCTGGCACAATTCGGCAGCCTTCTATCAATGTTGAAGTGTATACGC	792
750	Qy	AGCTTTTCTGTGATAGAGTCTCACTGTCTACCCAGGCTGGAGGCGAG	797
793	Db	TGTCCTCCATGCAGCACAAACCAACCGCCCAACCCAACTCTGAAGGCGCAG	840

```

RESULT 10
US-09-981-353-79
; Sequence 79, Application US/09981353
; Patent No. US20020160382A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy W.
; APPLICANT: Jones, David A.
; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
; FILE REFERENCE: PA-0038 US
; CURRENT APPLICATION NUMBER: US/09/981,353
; CURRENT FILING DATE: 2001-10-11
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 79
; LENGTH: 1244
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020160382A1 2101663C81
US-09-981-353-79

```

Query Match	36.2%	Score	329.6;	DB	9;	Length	1244;
Best Local Similarity	64.3%	Pred. No.	2.5e-95;				
Matches	494;	Conservative	0;	Mismatches	274;	Indels	0;
						Gaps	0;

Qy	30	GAACATGTCGAGGCTCAGCTGGGATACCGCGAGCAACACGGTCTATTCTCACTGGAAGG	89
Db	124	GATATGGCAAGTCCAGACTGGGATATGATGACAAAAATGGTCTGAACAATGGAGCAA	183
Qy	90	ATTTTCCCTATTGCTGATGGTGATCAGCAATCTCCAATTGAGATTAAAAACAAAGAAGT	149
Db	184	GCTGTATCCATTGGCAATGGAATAAACAGCTGCCCTGTTGATATTAAAAACCAAGTGAAC	243
Qy	150	GAATATGACTCTTCCCTCCGACCACCTTAGTATCAAGTATGATGCCCAAGCTCAGCTAAAT	209
Db	244	CAAAATGACACCTCTCTGAACCTATTAGTGTCTCCTACAAACCCAGCCACGCCAAGA	303

Qy	210	CATCAGCAACAGCGCCAA	TCTCTCAATGTTGACTTTTGATGACACAGAGAA	CAAAATCAGT	269
Db	304	AAATATCAATGTGGGGCAT	TCTTCCATGTAAATTTTGAGGACAACAGATAACCGATCAGT		363
Qy	270	TCTCGGTGGTCTCTCTCACTCGGAAGCTACAGGTTACGGCAGGTTCACTTTCAC	TGGGG	329	
Db	364	GCTGAAGGTGGTCTTTCTCTGACAGCTACAGGCTCTTTT	CAGTTCCTATTTTTCAC	TGGGG	423
Qy	330	GTCGGCTGATCACCACGGCTCCGAGCACATAGTAGATGGAGTGAGCTATGCTG	SCAGAGCT	389	
Db	424	CAGTACAAATGAGCATGGTTCAGAAACATACAGTGGATGGAGTCAAA	TATTTCTGCCGAGCT	483	
Qy	390	CCATGTTGTTCCACTGGAATTCAGACAAATACCCAGCTTTTGTGTAGGCGAGCTCATG	AAACC	449	
Db	484	TCAGTAGCTCACTGGAAATCTGCAAAAGTACTCCAGCCCTTGTG	AAAGCTGSCCTCAAAGGC	543	
Qy	450	AGATGAGCTGGCTGCTTGGGAGTGTTTTACAGATTGGTGAACTAACTAA	TCCCACTGCA	509	
Db	544	TGATGGTTTGGCAGTTATTTGGTGTTTTGATGAAGGTTGGTGAGGCCAACCC	AAAGGCTGCA	603	
Qy	510	AAAGATTAATCAGACATTTGGGATTTCCATTAAGAAAGGTTAAACAAACTCG	ATTACAAA	569	
Db	604	GABAGTACTGATGCCCTCCAAAGCAATTAACCAAGGCAACCGAGCCCCAT	TACAAA	663	
Qy	570	TTTTGACCTATTTGCTCTGCTTCCACCATCTCTGGGACTACTGGA	CATATCTCTGGTTCTCT	629	
Db	664	TTTTGACCCCTCTACTCTCTCTTCCCTTCATCCCTGGATTTCTGGA	CTACTCCCTGGCTCTCT	723	
Qy	630	TACAGTTCACCTCTCTTTGAGAGTGTCACATGGATTGTTTTAAAGCAACCTAT	ATAACAT	689	
Db	724	GACTCATCTCTCTTTAAGAGAGTGTAACCTTGGATCACTCTAAGGAGAGCAT	UAGTGT	783	
Qy	690	CAGCTCTCAACAGCTGGCCAAATTTTCGCAGTCTCTCTGTGCACAGCGGGGGTGA	AGCAGC	749	
Db	784	CAGCTCAGAGCAGCTGGCACAAATCCCGAGCCCTCTATCAAA	TGTTGAAGTGATAAACC	843	
Qy	750	AGCTTTTCTGTGATAGAGTCTCACTCTGTCAACCCAGGCTGGAGGGCAG		797	
Db	844	TGTCCCATGAGCACACAAACCGGCCAACCCAACTCTGAAGGGCAG		891	

```

RESULT 11
US-10-235-994-23
; Sequence 23, Application US/10235994
; Publication No. US20030101002A1
; GENERAL INFORMATION:
; APPLICANT: Bartha, Gabor
; APPLICANT: Walker, Michael
; TITLE OF INVENTION: METHODS FOR ANALYZING GENE EXPRESSION PATTERNS
; FILE REFERENCE: ICYPT012
; CURRENT APPLICATION NUMBER: US/10/235,994
; CURRENT FILING DATE: 2002-09-04
; PRIOR APPLICATION NUMBER: US/10/003,608
; PRIOR FILING DATE: 2001-11-01
; PRIOR APPLICATION NUMBER: 60/245,081
; PRIOR FILING DATE: 2000-11-01
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 1244
; TYPE: DNA
; ORGANISM: Human
US-10-235-994-23

```

Query Match	36.2%	Score	329.6;	DB	15;	Length	1244;
Best Local Similarity	64.3%	Pred. No.	2.5e-95;				
Matches	494;	Conservative	0;	Mismatches	274;	Indels	0;
						Gaps	0;

[illegible]

Db 184 GCTGATCCCAATGCGAATGGAATAACCGAGTCCCTGTTGATATTAACACCGTGAAC 243
QY 150 GAAATATGACTCTCCCTCCGACCACTTAGTAGTCAAGTATGACCCCAAGCTCAGCTAAAT 209
Db 244 CAAACATGACACTCTCTGAAACCTATTAGTCTCTCTCAACCCGACCCACAGCCCAAAGA 303
QY 210 CATGAGCAACAGCGGCATTCCTTCAATGTTGACTTTGATGACACAGAGAAACAAATCAGT 269
Db 304 AATTATCAATGTGGGGCATTCCTTCCATGTAAATTTTGAGGCAACAGATAACCGATCAGT 363
QY 270 TCTGGTGTGTCTCTCACTGGAAGCTACAGGTTTACGGAGGTTTCACTTCACTGGGG 329
Db 364 GCTGAAAGGTGGTCTCTCTGACAGCTACAGGCTCTTTCACTTCCATTTTCACTGGGG 423
QY 330 GTCCGCTGATGACCAACCGGCTCCGAGCACATAGTAGAGTGAGCTATGCTGCGAGCT 389
Db 424 CAGTACAAATGAGCATGGTTTCAGAACATACAGTGGATGGATCAAAATATTCTGCCGAGCT 483
QY 390 CCATGTTGTTCACTGGAATTCAGACAAATACCCGAGCTTTGTTGAGGCGAGCTCATGAAC 449
Db 484 TCACGTAGCTCACTGGAATTTCTGAAAGTACTCCAGCCCTTGTGAAAGCTGCCTCAAAGC 543
QY 450 AGATGAGCTGGTGTCTTCCGAGTGTCTTACAGATTGGTGAACCTTAATTTCCCAACTGCA 509
Db 544 TGATGTTTGGCAGTTATTGGTGTGTTGATGAAGTTGGTGAAGCTGCCTCAAAGC 603
QY 510 AAAGATTACTGACACTTTGGATTCCATTAAGAAAGGGTAAACAAACTCGATTCAAAA 569
Db 604 GAAAGTACTTGTGCTCCAGCAATTAACCAAGGGCAACAGAGCCCAATTCACAAA 663
QY 570 TTTTGACCTATTGCTCTGCTTCCACATCCTGGGACTACTGACATATCCTGGTCTCT 629
Db 664 TTTTGACCCCTTACTCTCTCTTCCATCCCTGGATTCTTGACCTTACCTGGCTCTCT 723
QY 630 TACAGTTCCACCTCTTCTTGAGAGTGTCAATGGATTGTTTAAAGCAACCTATAAACAT 689
Db 724 GACTCACTCTCTTTATGAGAGTGTAACTTGATCACTGTAAAGAGCATCAGTGT 783
QY 690 CAGCTCTAACAGCTGGCCAAATTTCCAGTCTCTGTCGACAGCGAGGGTGAAGCAGC 749
Db 784 CAGCTCAGAGCAGCTGGCACAATTCGCGAGCTTCTATCAAAATGTTGAAGGTGATAACGC 843
QY 750 AGCTTTTCTGTATAGAGTCTCACTCTGTCAACCCAGCTGGAGGGCAG 797
Db 844 TGTCCCATGTCAGCAACAACCGCCCAACCCCAACCTCTGAAGGGCAG 891

RESULT 12
US-10-158-646-47
; Sequence 47, Application US/10158646
; Publication No. US20030073105A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy K.W.
; APPLICANT: Sornasse, Thierry
; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
; FILE REFERENCE: FA-0030-1 US
; CURRENT APPLICATION NUMBER: US/10/158,646
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: 60/295,239
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PERL Program
; SEQ ID NO 47
; LENGTH: 2826
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20030073105A1 239568.5
US-10-158-646-47

Query Match 36.2%; Score 329.6; DB 15; Length 2826;

Best Local Similarity 64.3%; Pred. No. 4,1e-95;
Matches 494; Conservative 0; Mismatches 274; Indels 0; Gaps 0;
QY 30 GACCATGTGAGGCTCAGCTGGGGATACCGCAGACACACGCTCTCTATTCTCTGAGGA 89
Db 1218 GATAATGGCAAGTCCAGACTGGGGATATGATGACAAAATGCTCTGAAACATGGGACAA 1277
QY 90 ATTTTTCCTATTGCTGATGGTGATCAGCAATCTCCAATTGAGATTAAACCAACAGAACT 149
Db 1278 GCTGTATCCCAATGCAATGGAATAAACGAGTCCCTGTTGATATTAACACAGTGAAC 1337
QY 150 GAAATATGACTCTTCCCTCCGACCACTTAGTAGTCAAGTATGACCCCAAGCTCAGCTAAAT 209
Db 1338 CAAACATGACACTCTCTGAAACCTATTAGTCTCTCTACAAACCCAGCCACAGCCAAAGA 1397
QY 210 CATGAGCAACAGCGGCATTCCTTCAATGTTGACTTTGATGACACAGAGAAACAAATCAGT 269
Db 1398 AATTATCAATGTGGGGCATTCCTTCCATGTAAATTTTGAGGCAACAGATAACCGATCAGT 1457
QY 270 TCTGGTGTGTCTCTCTCACTGGAAGCTACAGGTTTACCGCAGGTTTCACTTCACTGGGG 329
Db 1458 GCTGAAAGGTGGTCTCTCTCTGACAGCTACAGGCTCTTTCACTTCCATTTTCACTGGGG 1517
QY 330 GTCCGCTGATGACCAACCGGCTCCGAGCACATAGTAGAGTGAGCTATGCTGCGAGCT 389
Db 1518 CAGTACAAATGAGCATGGTTTCAGAACATACAGTGGATGGATCAAAATATTCTGCCGAGCT 1577
QY 390 CCATGTTGTTCACTGGAATTCAGACAAATACCCGAGCTTTGTTGAGGCGAGCTCATGAAC 449
Db 1578 TCACGTAGCTCACTGGAATTTCTGAAAGTACTCCAGCCCTTGTGAAAGCTGCCTCAAAGC 1637
QY 450 AGATGAGCTGGTGTCTTCCGAGTGTCTTACAGATTGGTGAACCTTAATTTCCCAACTGCA 509
Db 1638 TGATGTTTGGCAGTTATTGGTGTGTTGATGAAGTTGGTGAAGCTGCA 1697
QY 510 AAAGATTACTGACACTTTGGATTCCATTAAGAAAGGGTAAACAAACTCGATTCAAAA 569
Db 1698 GAAAGTACTTGTGCTCCAGCAATTAACCAAGGGCAACAGAGCCCAATTCACAAA 1757
QY 570 TTTTGACCTATTGCTCTCTCTCCACCATCCTGGGACTACTGACATATCCTGGTCTCT 629
Db 1758 TTTTGACCCCTTACTCTCTCTTCCCTTCCCTGGATTCTTGACCTTACCTGGCTCTCT 1817
QY 630 TACAGTTCCACCTCTTCTTGAGAGTGTCAATGGATTGTTTAAAGCAACCTATAAACAT 689
Db 1818 GACTCACTCTCTTTATGAGAGTGTAACTTGATCACTGTAAAGAGCATCAGTGT 1877
QY 690 CAGCTCTAACAGCTGGCCAAATTTCCAGTCTCTGTCGACAGCGAGGGTGAAGCAGC 749
Db 1878 CAGCTCAGAGCAGCTGGCACAATTCGCGAGCTTCTATCAAAATGTTGAAGGTGATAACGC 1937
QY 750 AGCTTTTCTGTATAGAGTCTCACTCTGTCAACCCAGCTGGAGGGCAG 797
Db 1938 TGTCCCATGTCAGCAACAACCGCCCAACCCCAACCTCTGAAGGGCAG 1985

RESULT 13
US-09-917-800A-1675
; Sequence 1675, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Elashoff, Michael
; APPLICANT: Elashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31

Db 375 TGAACAC---CAATATGGGATTTTGGAAAAGCTGTGCAGCACCCAGATGGCTGGCT 431
Qy 463 GCTCTGGGAGTGTCTTTTACAGATGGTGAACCTAATTCCTCAACTGCAAAAGATTACTGAC 522
Db 432 GTTTTGGGTATTTTGTGAATGGACCTGCTCACAAGGCTTCAGAAAATCACTGAA 491
Qy 523 ACTTTGGATTCATTAAGAAAGGTAAACAACTCGATTTCACAAATTTTGCACCTATTG 582
Db 492 GCACGTGATTCATTAACAAAGGGGAAACGTGCAGCCCTTTTGTAACTTTGATCCTTGC 551
Qy 583 TCTCTGCTTCCACCATCTCTGGGACTACTGACATATCTCTGTTCTTACAGTTCCACCT 642
Db 552 TCCCTCTCTTCTGGAAACTTGGACTACTGACATATCTCTGCTCTCTGACCACTCGCCC 611
Qy 643 CTTCTTGAGAGTGTACATGGATTGTTTAAAGCAACCTATATAACATCAGCTCTCAACAG 702
Db 612 CTGCTGAATGTGTGACCTGGATAGTGCTCAAGGAACCCATTACTGTGACGAGTGAGCAG 671
Qy 703 CTGGCCAAATTCGAGTCTCTCTGTCACAGCGGAGGTGAAGCAGCAGCTTTCTGTGA 762
Db 672 ATGTCTCATTTCCGTAACCTGAACCTTCAATTCGAGGGGGAGGCTGAAGAACTGATGGTG 731
Qy 763 TAGAGTCTCACTGTGCACCCAGGCTGGAGGCGAG 797
Db 732 GACAACTGGGCTCAGCTAGCGCTGAAGAACAG 766

RESULT 15

US-09-988-292-17
; Sequence 17, Application US/09988292
; Publication No. US20020086314A1
; GENERAL INFORMATION:
; APPLICANT: Yu, Guo-Liang
; Rosen, Craig
; TITLE OF INVENTION: Colon Specific Genes and Proteins
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi,
; Stewart & Olstein
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: NJ
; COUNTRY: USA
; ZIP: 07068-1739
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/988,292
; FILING DATE: 19-Nov-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/224,110
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferraro, Gregory D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-435
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 605 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..603

FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 1..603
; SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-09-988-292-17
Query Match 31.3%; Score 284.8; DB 13; Length 605;
Best Local Similarity 67.2%; Pred. No. 4.9e-81;
Matches 403; Conservative 0; Mismatches 197; Indels 0; Gaps 0;
Qy 128 TTGAGATTAAACCAAAAGGTGAATATGATCTTCCCTCCGACCACTAGTATCAAGT 187
Db 2 TTGATATTAAACCAAGTGAACCAACCATGACACCTCTCTGAAACCTATTAGTGTCTCT 61
Qy 188 ATGACCCCAAGCTCAGCTAAATCATCAGCAACAGCGGCGCATTCCTTCAATGTTGACTTTG 247
Db 62 ACACCCCAAGCTCAGCTAAATCATCAGCAACAGCGGCGCATTCCTTCAATGTTGACTTTG 121
Qy 248 ATGACACAGAGAACAAATCAGTTCTGCGTGGTCTCTCTCACTGGAAGTACAGGTTAC 307
Db 122 AGGACACCAATTAACCGATCAGTGTGAAAGGTGGTCTTCTCTGACAGCTACAGGCTCT 181
Qy 308 GGCAGGTTCACTTCTGCGGCTCGCTGATGATGACACCGCTCCGAGCAGCATAGTAGATG 367
Db 182 TTCAGTTCCATTTTCACTGGGCGAGTACAAATGAGCATGGTTTCAGAACATACAGTGGATG 241
Qy 368 GAGTGAGCTATGTCGACAGAGCTCCATGTTGTTTCACTGGAATTCAGACAAATACCCAGCT 427
Db 242 GAGTCAATATTCTGCCGAGCTTCACTGCTCACTGGAAATTCGCAAGTACTCCAGCC 301
Qy 428 TTGTTGAGGCGCTCATGAACCAAGATGGAATGCTGCTCTGTTGGAGTGTGTTTACAGATTG 487
Db 302 TTGCTGAAGTGCCTCAAGGCTGATGGTTGGCAGTTATTTGTTGTTGATGAAGTTG 361
Qy 488 GTGAACCTAATTCCTCAACTGCAAAAGATTAATGACACTTTGGATTCCATTAAAGAAAAG 547
Db 362 GTGAGGCCAACCCAAAGCTGCAGAAAGTACTTGTATGCCCTCCAGCAATTAAGAACCAAG 421
Qy 548 GTAAACAACTCGATTTCACAAATTTTGACCTATTGTTCTCTGCTTCCACCATCTCTGGGACT 607
Db 422 GCAACGAGCCCCCAATTCACAAATTTTGACCCCTCTACTCTCTTCTCTATCCCTGGATT 481
Qy 608 ACTGGACATATCCTGTTCTCTTACAGTTCCACCTCTTCTTGTGAGAGTGTACATGGATTG 667
Db 482 TCTGGACCTACCTGCTCTCTGACTCATCTCTCTTTATGAGAGTGTAACTTGGATCA 541
Qy 668 TTTTAAAGCAACCTATAACATCAGCTCTCAACAGCTGGCCAAATTTTCGAGTCTCTCTGT 727
Db 542 TCTGTAAGGAGAGCATCAGTGTCAAGTTTCAGAGAGTTGGCACAAATTCGGAGCCTTCTAT 601

Search completed: September 14, 2004, 01:34:58
Job time : 502 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 2004, 20:55:55 ; Search time 424 Seconds
(without alignments)
9127.614 Million cell updates/sec

Title: US-10-069-434-4
Perfect score: 911
Sequence: 1 cggattcggctcgagttccc.....cccaaatccagggggaaaaa 911

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq 29Jan04:.*
1: Geneseqn1980s:.*
2: Geneseqn1990s:.*
3: Geneseqn2000s:.*
4: Geneseqn2001as:.*
5: Geneseqn2001bs:.*
6: Geneseqn2002s:.*
7: Geneseqn2003as:.*
8: Geneseqn2003bs:.*
9: Geneseqn2003cs:.*
10: Geneseqn2004s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	911	100.0	911	6	ABA97691 Human lya
2	884.8	97.1	2094	7	ABX34690 Human mld
3	759	83.3	1248	4	AAS30989 Human dia
4	745.8	81.9	1023	6	AAD41388 Human NZM
5	740.6	81.3	828	7	ABT33350 NOVX DNA
6	728.8	80.0	833	7	ABT33351 NOVX DNA
7	493.2	54.1	688	3	Aaf22342 Human sec
8	444.4	48.8	462	5	AAS68431 DNA encod
9	444.4	48.8	496	8	ACH26864 Human adu
10	421	46.2	462	4	AAI92460 Human pol
11	331.2	36.4	2785	6	AAU43637 Human car
12	329.6	36.2	1233	4	Aah33294 Human col
13	329.6	36.2	1244	8	ADAI0961 Human cdn
14	329.6	36.2	2826	9	AAD59160 Human car
15	289.4	31.8	1459	6	ABK63768 Rat seque
16	289.4	31.8	1459	9	ADBS8362 Toxicity-
17	284.8	31.3	605	2	AAT45889 Human col
18	284.8	31.3	605	2	AAV16677 Polynucle
19	284.8	31.3	605	6	ABK15465 Human col
20	284.8	31.3	605	6	ABS52546 Human col
21	280	30.7	783	2	AAV47615 Nucleotid
22	280	30.7	864	2	AAT73088 DNA for f
23	280	30.7	1759	8	ADAI0998 Human cdn

24	280	30.7	1759	8	ACH04029 Human cdn
25	278.4	30.6	1523	6	ABL64977 lung canc
26	278.4	30.6	1523	6	ABL62488 Colon ade
27	278.4	30.6	1523	6	ABL68606 Kidney ca
28	278.4	30.6	1523	6	ABN97349 Gene #384
29	278.4	30.6	1723	3	AAC98162 Human col
30	278.4	30.6	1755	6	AAS94944 Human DNA
31	278.4	30.6	1755	9	AAD59159 Human car
32	278.4	30.6	2391	6	ABK35712 cDNA sequ
33	266.8	29.3	1266	6	ABA92275 Streptavi
34	266.8	29.3	1765	4	AH57384 Human ske
35	266.8	29.3	2286	6	ABO93292 Human cdn
36	266.2	28.2	1067	6	ABI99672 Mouse isc
37	255	28.0	988	6	ABK3769 Rat seque
38	255	28.0	988	9	ADBS8363 Toxicity-
39	247	27.1	247	6	ABQ72494 Human MDD
40	213.6	23.4	702	6	ABT09204 Phase-1 R
41	200.4	22.0	791	5	AAS68432 DNA encod
42	198.8	21.8	1825	6	ABZ11737 Human pol
43	197.8	21.7	507	8	ADAI0984 Human cdn
44	197.6	21.7	1201	6	ABK63770 Rat seque
45	197.6	21.7	1201	9	ADBS8364 Toxicity-

ALIGNMENTS

RESULT 1
ABA97691
ID ABA97691 standard; DNA; 911 BP.
XX
AC ABA97691;
XX
DT 18-JUN-2002 (first entry)
XX
DE Human lyase HLYA-1 encoding cDNA.
XX
KW Cytostatic; anticonvulsant; cerebroprotective; nootropic; virucide;
KW neuroprotective; antibacterial; antidiabetic; antiinflammatory; antitumor;
KW ophthalmologic; hypotensive; immunosuppressive; dermatological;
KW nephrologic; antithyroid; thyromimetic; osteopathic; antipsoriatic;
KW antileukic; fungicide; antiparasitic; protozoacide; tranquiliser; cancer;
KW neuroleptic; diagnosis; treatment; immunological disorder; AIDS; allergy;
KW acquired immunodeficiency syndrome; asthma; HLYA; infection; anaemia;
KW Crohn's disease; multiple sclerosis; atherosclerosis; osteoporosis;
KW rheumatoid arthritis; Alzheimer's; Parkinson's disease; epilepsy; stroke;
KW muscular dystrophy; Down's syndrome; myasthenia gravis; glaucoma;
KW transgenic; gene therapy; drug screening; human lyase; gene; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
CDS 34..762
FT /*tag= a
FT /product= "HLYA-1"
FT /note= "Human lyase 1"
XX
WO200200840-A2.
XX
PD 03-JAN-2002.
XX
PF 13-JUN-2001; 2001WO-US019166.
XX
PR 23-JUN-2000; 2000US-0213383P.
PR 30-JUN-2000; 2000US-0215544P.
PR 04-AUG-2000; 2000US-0222818P.
XX
(INCY-) INCYTE GENOMICS INC.
XX
PI Thornton M, Ramkumar J, Tribouley CM, Yue H, Nguyen DB, Yao MG;
PI Patterson C, Gandhi AR, Burford N, Thangavelu K, Baughn MR;
XX
DR WPI; 2002-139910/18.

DR P-PSDB; ABB08900.
XX New isolated human lyase polypeptide for diagnosing, treating and
PT preventing e.g. glaucoma, ocular hypertension, stroke, asthma, or gout.
XX Claim 11; Page 99-100; 101pp; English.
XX
CC The present sequence represents a cDNA encoding a human lyase polypeptide
CC (HLXA-1) given in ABB08900. The specification describes an isolated HLXA
CC polypeptide or a nucleic acid that encodes it. The invention has
CC cytotostic, nootropic, anticonvulsant, cerebroprotective, virucide,
CC neuroprotective, anti-HIV, antiparkinsonian, antibacterial, antidiabetic,
CC antiinflammatory, ophthalmological, hypotensive, antiallergic, antitumor,
CC antianemic, antisthmatic, antiatherosclerotic, immunosuppressive,
CC dermatological, nephrotropic, thymomimetic, osteopathic, antipsoriatic,
CC antirheumatic, antiarthritic, dermatological, fungicide, antiparasitic,
CC protozoacide, tranquilizer and neuroleptic applications. The protein of
CC the invention may be used to screen for potential HLXA agonists or
CC antagonists; detect the presence of HLXA-associated disorders; assess the
CC toxicity of a test compound. The HLXA proteins and polynucleotides are
CC useful in diagnosis, treatment and prevention of immunological disorders
CC e.g. AIDS, allergy, anaemia, asthma, infection, Crohn's disease, multiple
CC sclerosis, atherosclerosis, rheumatoid arthritis, osteoporosis; cancer;
CC neurological disorder e.g. Alzheimer's and Parkinson's disease, epilepsy,
CC stroke, muscular dystrophy, Down's syndrome, myasthenia gravis; glaucoma.
CC HLXA polynucleotides are used for creating humanised/transgenic animals
CC to model human diseases; somatic or germline gene therapy; for generating
CC hybridisation probes for e.g. gene mapping; detecting differences in
CC chromosomal location due to e.g. translocation; generating a transcript
CC image of a tissue/cell type. Antibodies which bind to the HLXA proteins
CC are used for diagnosis of HLXA-associated disorders or monitoring
CC patients being treated with HLXA or agonists, antagonists or inhibitors
CC of HLXA and for assessing toxicity of a test compound
XX
SQ Sequence 911 BP; 238 A; 235 C; 205 G; 233 T; 0 U; 0 Other;
Query Match 100.0%; Score 911; DB 6; Length 911;
Best Local Similarity 100.0%; Pred. No. 3.7e-286;
Matches 911; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CGGAATTCGGCTCGAGTTCACCCCGAGGACCACATGTCGAGGTCAGCTGGGATACCCG 60
Db 1 CGGAATTCGGCTCGAGTTCACCCCGAGGACCACATGTCGAGGTCAGCTGGGATACCCG 60
QY 61 GAGCACAAACGGTCTTACCTGGAAGAAATTTTCCTATTGCTGATGATGATCAGCAA 120
Db 61 GAGCACAAACGGTCTTACCTGGAAGAAATTTTCCTATTGCTGATGATGATCAGCAA 120
QY 121 TCTCCAAATTCAGATTAAACCAAGAGTGAATATGACTCTTCCTCCGACCACTTAGT 180
Db 121 TCTCCAAATTCAGATTAAACCAAGAGTGAATATGACTCTTCCTCCGACCACTTAGT 180
QY 181 ATCAAGTATGACCAAGCTCAGCTAAATCATCAGCAACACGCGCCATTCCTTCAATGTT 240
Db 181 ATCAAGTATGACCAAGCTCAGCTAAATCATCAGCAACACGCGCCATTCCTTCAATGTT 240
QY 241 GACTTTGATGACACAGAGAACAAATCAGTTCTCGGTGGTCTCTCACTGGAAGCTAC 300
Db 241 GACTTTGATGACACAGAGAACAAATCAGTTCTCGGTGGTCTCTCACTGGAAGCTAC 300
QY 301 AGGTACGGCAGGTTTCACCTTCACTGGGGTTCGCTGATGACCAAGCTCCGAGCACATA 360
Db 301 AGGTACGGCAGGTTTCACCTTCACTGGGGTTCGCTGATGACCAAGCTCCGAGCACATA 360
QY 361 GTAGATGGAGTGAAGTATGCTGACAGAGCTCCATGTTGTTCACTGGAATTCAGACAAATAC 420
Db 361 GTAGATGGAGTGAAGTATGCTGACAGAGCTCCATGTTGTTCACTGGAATTCAGACAAATAC 420
QY 421 CCCAGCTTTGTTGAGGAGCTCATGAACACAGATGAGTGGCTGCTTCTGGGAGTGTTTTA 480
Db 421 CCCAGCTTTGTTGAGGAGCTCATGAACACAGATGAGTGGCTGCTTCTGGGAGTGTTTTA 480
QY 481 CAGATTGTTGAACTAATTCCTCACTGCAAAAGATTACTGACACTTTTGGATTTCATTAAA 540

Db 481 CAGATTGTTGAACTAATTCCTCACTGCAAAAGATTACTGACACTTTTGGATTTCATTAAA 540
QY 541 GAAAAGGGTAAACAAACTCGATTCAAAATTTTGACCTATTGTTCTCTGTTCCACCATCC 600
Db 541 GAAAAGGGTAAACAAACTCGATTCAAAATTTTGACCTATTGTTCTCTGTTCCACCATCC 600
QY 601 TGGGACTACTGGACATATCTCGTTCTCTTACAGTTCCACCTCTTCTTGGAGAGTGTCA 660
Db 601 TGGGACTACTGGACATATCTCGTTCTCTTACAGTTCCACCTCTTCTTGGAGAGTGTCA 660
QY 661 TGGATTGTTTAAAGCAACCTATAAATCATCAGCTCTCAACAGCTGGCCAAATTTGCA 720
Db 661 TGGATTGTTTAAAGCAACCTATAAATCATCAGCTCTCAACAGCTGGCCAAATTTGCA 720
QY 721 CTCCTGTCACAGCGAGGGTGAAGCAGCAGCTTTTCTGTGATAGAGTCTCACTCTGTCA 780
Db 721 CTCCTGTCACAGCGAGGGTGAAGCAGCAGCTTTTCTGTGATAGAGTCTCACTCTGTCA 780
QY 781 CCCAGCTGGAGGGCAGTGTACATCTTGGCTTAATTCAGCTCCAACTCTCGGACTCA 840
Db 781 CCCAGCTGGAGGGCAGTGTACATCTTGGCTTAATTCAGCTCCAACTCTCGGACTCA 840
QY 841 AGTGATCTCTCCACCTCAGCTCCAGAGTCTCGAGCTCCAGATGATGATGATGATGATG 900
Db 841 AGTGATCTCTCCACCTCAGCTCCAGAGTCTCGAGCTCCAGATGATGATGATGATGATG 900
QY 901 CAGGGGGGAAAA 911
Db 901 CAGGGGGGAAAA 911
RESULT 2
ABX34690
ID ABX34690 standard; cDNA; 2094 BP.
XX AC ABX34690;
XX AC ABX34690;
DT 13-FEB-2003 (first entry)
XX Human mddt cDNA SEQ ID 251.
XX MDDT; human; disease detection and treatment molecule polypeptide;
KW anti-inflammatory; immunosuppressive; osteopathic; cytostatic; anti-HIV;
KW haemostatic; nephrotropic; antianemic; antipsoriatic; hepatotropic;
KW gene therapy; protein replacement therapy; cell proliferative disorder;
KW cancer; adenocarcinoma; leukaemia; lymphoma; melanoma; sarcoma;
KW anaemia; Crohn's disease; acquired immunodeficiency syndrome; AIDS;
KW Goodpasture's syndrome; inflammation; osteoporosis; thrombocytopaenia;
KW psoriasis; hepatitis; gene; ss.
XX Homo sapiens.
XX WO20279449-A2.
XX 10-OCT-2002.
XX 27-MAR-2002; 2002WO-US009944.
XX 28-MAR-2001; 2001US-0279619P.
XX 29-MAR-2001; 2001US-0280067P.
XX 29-MAR-2001; 2001US-0280068P.
XX 16-MAY-2001; 2001US-0291280P.
XX 17-MAY-2001; 2001US-0291829P.
XX 17-MAY-2001; 2001US-0291849P.
XX 19-JUN-2001; 2001US-0299428P.
XX 20-JUN-2001; 2001US-0299776P.
XX 20-JUN-2001; 2001US-0300001P.
XX (INCY-) INCYTE GENOMICS INC.
XX Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;
XX Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Anshey SR;
PI

PI Daugherty SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;
PI Peralta CH, David MR, Lewis SA, Chen AJ, Panzer SR, Harris B;
PI Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;
XX WPI; 2003-058431/05.
DR P-PSDB; ABU11700.
XX
XX
PT New purified disease detection and treatment molecule proteins and
PT polynucleotides, useful for diagnosing, treating or preventing cancers
PT (e.g. leukemia or sarcoma), anemia, Crohn's disease, AIDS, osteoporosis
PT or hepatitis.
PS
PS Claim 1; SEQ ID NO 251; 339pp + Sequence Listing; English.
XX
XX This invention describes a novel disease detection and treatment molecule
CC polypeptide (MDTR) which has anti-inflammatory, immunosuppressive,
CC osteopathic, cytostatic, anti-HIV, haemostatic, nephrotropic,
CC antianemic, antipsoriatic and hepatotropic activity. The polynucleotides
CC and the polypeptides of the invention can be used for gene therapy,
CC protein replacement therapy and are useful for treating a variety of
CC diseases or conditions. These polypeptides or polynucleotides are
CC particularly useful for diagnosing, treating or preventing cell
CC proliferative disorders (e.g. cancers including adenocarcinoma,
CC leukemia, lymphoma, melanoma, myeloma or sarcoma), anaemia, Crohn's
CC disease, acquired immunodeficiency syndrome (AIDS), Goodpasture's
CC syndromes, inflammation, osteoporosis, thrombocytopaenia, psoriasis or
CC hepatitis. ABX344435 encode the MDR polypeptides represented in
CC ABU11450-ABU11845, described in the disclosure of the invention. NOTE:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format from WIDO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2094 BP; 555 A; 500 C; 480 G; 559 T; 0 U; 0 Other;

Query Match
Best Local Similarity 97.1%; Score 884.8; DB 7; Length 2094;
Matches 908; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 1 CGGAATTCGGCTCGAGTTCCACCCGAGGACATGTCGAGGCTCAGTGGGATACCGC 60
DB |||||||
15 CGGAATTCGGCTCGAGTTCCACCCGAGGACATGTCGAGGCTCAGTGGGATACCGC 74
QY 61 GAGCAACGGTCCCTATTCACCTGAGGAGGATTTTCCCTATTCCTGATGATCAGCAA 120
DB |||||||
75 GAGCAACGGTCCCTATTCACCTGAGGAGGATTTTCCCTATTCCTGATGATCAGCAA 134
QY 121 TCTCAAATGAGATTAAACCAAGAGTGAATATGACTTCTCCCTCCGACCACTTAGT 180
DB |||||||
135 TCTCAAATGAGATTAAACCAAGAGTGAATATGACTTCTCCCTCCGACCACTTAGT 194
QY 181 ATCAAGTATGACCAAGCTCAGCTAAATCATCAGCAAGCGGCGCATTCCTTCAATGTT 240
DB |||||||
195 ATCAAGTATGACCAAGCTCAGCTAAATCATCAGCAAGCGGCGCATTCCTTCAATGTT 254
QY 241 GACTTTGATGACAGAGAACAAATCAGTTCCTCGGTGGTGGTCTCTCACTGGAAGCTAC 300
DB |||||||
255 GACTTTGATGACAGAGAACAAATCAGTTCCTCGGTGGTGGTCTCTCACTGGAAGCTAC 314
QY 301 AGGTTACGGCAGGTTCACTTCACTGGGGGTCGGCTGATGACACGGCTCCGAGCACATA 360
DB |||||||
315 AGGTTACGGCAGGTTCACTTCACTGGGGGTCGGCTGATGACACGGCTCCGAGCACATA 374
QY 361 GTAGATGAGTGAAGTATCTGCTGAGAGCTCCATGTTGTTTCACTGGAATTCAGCAAAATAC 420
DB |||||||
375 GTAGATGAGTGAAGTATCTGCTGAGAGCTCCATGTTGTTTCACTGGAATTCAGCAAAATAC 434
QY 421 CCAGGTTTGTGAGGACGCTGATGAACAGATGAGCTGCTGCTGGAGTGTTTGA 480
DB |||||||
435 CCAGGTTTGTGAGGACGCTGATGAACAGATGAGCTGCTGCTGGAGTGTTTGA 494
QY 481 CAGATTGGTGAACCTAAATTCACCACTGCAAAAGATTACTGACACTTTGATTCCATTA 540
DB |||||||
495 CAGATTGGTGAACCTAAATTCACCACTGCAAAAGATTACTGACACTTTGATTCCATTA 554

QY 541 GAAAAGGGTAACAAACATCGATTCAAAATTTTGACCTATTGCTCTGCTTCCACCATCC 600
DB |||||||
555 GAAAAGGGTAACAAACATCGATTCAAAATTTTGACCTATTGCTCTGCTTCCACCATCC 614
QY 601 TGGGACTACTGGACATATCTCGTTCTCTTACAGTTCCACCTCTTCTTGGAGAGTGTACA 660
DB |||||||
615 TGGGACTACTGGACATATCTCGTTCTCTTACAGTTCCACCTCTTCTTGGAGAGTGTACA 674
QY 661 TGGATTGTTTAAAGCAACCTATAAACAATCAGCTCTCAACAGCTGGCCAAATTTGCGAGT 720
DB |||||||
675 TGGATTGTTTAAAGCAACCTATAAACAATCAGCTCTCAACAGCTGGCCAAATTTGCGAGT 734
QY 721 CTCCTGTGCACAGCGGAGGTGAAGCAGAGCTTTTCTGTGATAGAGTCTACTCTGTCA 780
DB |||||||
735 CTCCTGTGCACAGCGGAGGTGAAGCAGAGCTTTTCTGTGATAGAGTCTACTCTGTCA 794
QY 781 CCAGGCTGGAGGAGGTGATCAATCTTGGCTAATTCAGGCTCCAACTCTCGACTCA 840
DB |||||||
795 CCAGGCTGGAGGAGGTGATCAATCTTGGCTAATTCAGGCTCCAACTCTCGACTCA 853
QY 841 AGTGATCTCTCCACCTCAGCTCCAGAGTCTCAGAGTCTCAGAGTGGCATG-ACITTTCCCAAATG 899
DB |||||||
854 AGTGATCTCTCCACCTCAGCTCCAGAGTCTCAGAGTGGCATGAGAACTTTTCCCAAATG 913
QY 900 CCAGGCGGAAAA 911
DB |||||||
914 CCAGGCGGAAA 925

RESULT 3
AAS30989
ID AAS30989 standard; cDNA; 1248 BP.
XX
AC AAS30989;
XX
DT 04-DEC-2001 (first entry)
XX
DE Human diagnostic and therapeutic polynucleotide (DITHP) #4.
XX
KW Human; receptor; diagnostic; therapeutic; gene therapy; vaccine;
KW cell proliferative disorder; Crohn's disease; lymphoma; leukaemia;
KW acquired immune deficiency syndrome; AIDS; autoimmune disorder;
KW respiratory disorder; ss.
OS Homo sapiens.
XX
PN WO200162927-A2.
XX
PD 30-AUG-2001.
XX
PF 21-FEB-2001; 2001WO-US006059.
XX
PR 24-FEB-2000; 2000US-0184693P.
PR 24-FEB-2000; 2000US-0184697P.
PR 24-FEB-2000; 2000US-0184698P.
PR 24-FEB-2000; 2000US-0184768P.
PR 24-FEB-2000; 2000US-0184769P.
PR 24-FEB-2000; 2000US-0184770P.
PR 24-FEB-2000; 2000US-0184771P.
PR 24-FEB-2000; 2000US-0184772P.
PR 24-FEB-2000; 2000US-0184773P.
PR 24-FEB-2000; 2000US-0184774P.
PR 24-FEB-2000; 2000US-0184776P.
PR 24-FEB-2000; 2000US-0184777P.
PR 24-FEB-2000; 2000US-0184797P.
PR 24-FEB-2000; 2000US-0184813P.
PR 24-FEB-2000; 2000US-0184837P.
PR 24-FEB-2000; 2000US-0184841P.
PR 24-FEB-2000; 2000US-0185213P.
PR 24-FEB-2000; 2000US-0185216P.
PR 12-MAY-2000; 2000US-0203785P.
PR 15-MAY-2000; 2000US-0204226P.

PR 16-MAY-2000; 2000US-0204525P.
PR 16-MAY-2000; 2000US-0204821P.
PR 16-MAY-2000; 2000US-0204908P.
PR 16-MAY-2000; 2000US-0205232P.
PR 17-MAY-2000; 2000US-0204815P.
PR 17-MAY-2000; 2000US-0204863P.
PR 17-MAY-2000; 2000US-0205221P.
PR 17-MAY-2000; 2000US-0205285P.
PR 17-MAY-2000; 2000US-0205286P.
PR 17-MAY-2000; 2000US-0205287P.
PR 17-MAY-2000; 2000US-0205323P.
PR 17-MAY-2000; 2000US-0205324P.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Panzer SR, Spiro PA, Banville SC, Shah P, Chalup MS, Chang SC;
PI Chen A, D'sa SA, Anshey S, Dahl CR, Dam TC, Daniels SE, Dufour GR;
PI Flores V, Fong WT, Greenawalt LB, Hillman JL, Jones AL, Liu TF;
PI Roseberry AM, Rosen BH, Russo ED, Stockdreher TK, Daffo A;
PI Wright RJ, Yap PE, Yu JY, Bradley DU, Bratcher SR, Chen W;
PI Cohen HJ, Hodgson DW, Lincoln SE, Jackson S;
XX
XX WPI; 2001-502867/55.
DR P-PSDB; AAU19418.
XX
XX Polynucleotides encoding diagnostic and therapeutic proteins, e.g.
PT enzymes, hormones and receptors, useful in diagnostics and therapeutics.
PT
XX
XX Claim 1; Page 297; 522pp; English.
XX
XX The invention relates to polynucleotides (I) encoding diagnostic and
CC therapeutic (DITHP) polypeptides (II), which include e.g. enzymes, and
CC proteins involved in growth and development and receptors. (I) and (II)
CC may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate DITHP expression. For example, (I) and (II)
CC may be used to treat disorders associated with decreased polypeptide
CC expression by rectifying mutations or deletions in a patient's genome,
CC that affect the activity of the DITHPs, by expressing inactive proteins
CC or supplementing the patient's own production of them. (I) and (II) may
CC be used to treat diseases, for example, cell proliferative disorder,
CC Crohn's disease, acquired immune deficiency syndrome (AIDS), lymphoma,
CC leukaemia, autoimmune disorders, and respiratory disorders. Additionally,
CC (I) may be used to produce the DITHPs by inserting the nucleic acids
CC into a host cell and culturing the cell to express the protein. (I) and
CC its complementary sequences may also be used as DNA probes in diagnostic
CC assays to detect and quantitate the presence of similar nucleic acids in
CC samples, and therefore which patients may be in need of restorative
CC therapy. (II) may also be used as antigens in the production of
CC antibodies against DITHPs and in assays to identify modulators of DITHP
CC expression and activity. The anti-DITHP antibodies and antagonists may
CC also be used to down regulate expression and activity. The anti-DITHP
CC antibodies may also be used as diagnostic agents for detecting the
CC presence of DITHPs in samples (e.g. by enzyme linked immunosorbant assay
CC (ELISA)). AAS30986-AAS31196 represent human diagnostic and therapeutic
CC (DITHP) polynucleotides of the invention
XX
XX Sequence 1248 BP; 326 A; 294 C; 283 G; 343 T; 0 U; 2 Other;
SQ

Query Match 83.3%; Score 759; DB 4; Length 1248;
Best Local Similarity 100.0%; Pred. No. 1.8e-236;
Matches 759; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CGGAATTCGGCTCGAGTTCCACCCCGGAGGACCATGTCGAGGCTCAGCTGGGATACCCG 60
DB 15 CGGAATTCGGCTCGAGTTCCACCCCGGAGGACCATGTCGAGGCTCAGCTGGGATACCCG 74
QY 61 GAGCACAAACGGTCTCTATTCTACTGGAAGAAATTTTCCCTATTGCTGATGTCATGACGAA 120
DB 75 GAGCACAAACGGTCTCTATTCTACTGGAAGAAATTTTCCCTATTGCTGATGTCATGACGAA 134
QY 121 TCTCCCAATTGAGATTAAACCAAGAGTGAATATGACTCTTCCCTCCGACCACTTAGT 180
DB 135 TCTCCCAATTGAGATTAAACCAAGAGTGAATATGACTCTTCCCTCCGACCACTTAGT 194

QY 181 ATCAAGTATACCCCAAGCTCAGCTAAATATCATCAGCAACAGCGGCATTCTCTCAATGTT 240
DB 195 ATCAAGTATACCCCAAGCTCAGCTAAATATCATCAGCAACAGCGGCATTCTCTCAATGTT 254
QY 241 GACTTTGATGACACAGAGAAATAATCAGTTCTGCGTGGTGGTCTCTCACTGGAAGCTAC 300
DB 255 GACTTTGATGACACAGAGAAATAATCAGTTCTGCGTGGTGGTCTCTCACTGGAAGCTAC 314
QY 301 AGGTTACGGCAGGTTCACTTCACTGGGGGTCCTGATGACCAAGGCTCCGAGCACATA 360
DB 315 AGGTTACGGCAGGTTCACTTCACTGGGGGTCCTGATGACCAAGGCTCCGAGCACATA 374
QY 361 GTAGATGAGTGAGCTATGCTGCAGAGCTCCATGTTGTTTCACTGGAATTCAGACAAATAC 420
DB 375 GTAGATGAGTGAGCTATGCTGCAGAGCTCCATGTTGTTTCACTGGAATTCAGACAAATAC 434
QY 421 CCCAGCTTTGTTGAGGCAGCTCATGAACAGATGGCTGTCTTGGGAGTGTGTTT 480
DB 435 CCCAGCTTTGTTGAGGCAGCTCATGAACAGATGGCTGTCTTGGGAGTGTGTTT 494
QY 481 CAGATTGGTGAACCTTAATCCCAACTGCAAAAGATTACTGACACTTTGGATTCCATTAA 540
DB 495 CAGATTGGTGAACCTTAATCCCAACTGCAAAAGATTACTGACACTTTGGATTCCATTAA 554
QY 541 GAAAGGGTAAACAAACTCGATTCAAAATTTTGACCTATTGTTCTGCTTCCACCATCC 600
DB 555 GAAAGGGTAAACAAACTCGATTCAAAATTTTGACCTATTGTTCTGCTTCCACCATCC 614
QY 601 TGGGACTACTGGACATATCCTGGTCTCTTACAGTTCACCTCTTCTTGAGAGTGTGACA 660
DB 615 TGGGACTACTGGACATATCCTGGTCTCTTACAGTTCACCTCTTCTTGAGAGTGTGACA 674
QY 661 TGGATTGTTTAAAGCAACCTATAACATCAGCTCTCAACAGCTGCCCCAAATTCGCAGT 720
DB 675 TGGATTGTTTAAAGCAACCTATAACATCAGCTCTCAACAGCTGCCCCAAATTCGCAGT 734
QY 721 CTCTGTGACAGCGGAGGCTGAAGCAGCAGCTTTTCTG 759
DB 735 CTCTGTGACAGCGGAGGCTGAAGCAGCAGCTTTTCTG 773
RESULT 4
AAD41388
ID AAD41388 standard; cDNA; 1023 BP.
XX AAD41388;
XX 30-OCT-2002 (first entry)
XX Human NZMS-1 cDNA.
XX Human; enzyme; NZMS-1; cell proliferative disorder; hepatitis; cancer;
KW psoriasis; leukaemia; autoimmune disorder; diabetes; arteriosclerosis;
KW acquired immune deficiency syndrome; cardiovascular disorder; anorexia;
KW asthma; hypertension; gastrointestinal disorder; reproductive disorder;
KW gastritis; neurological disorder; epilepsy; pulmonary disorder; AIDS;
KW dementia; embolism; gene therapy; eye disorder; transgenic; gene; ss.
XX Homo sapiens.
XX Key Location/Qualifiers
CDS 18..806
FT /*tag= a
FT /product= "NZMS-1 protein"
FT /transl_except= (pos:240..290, aa:Glu-leu)
XX WO200246385-A2.
XX 13-JUN-2002.
XX 04-DEC-2001; 2001WO-US047432.
XX

07-DEC-2000; 2000US-0251824P.
 08-DEC-2000; 2000US-0254312P.
 14-DEC-2000; 2000US-0255773P.
 15-DEC-2000; 2000US-0255940P.
 15-DEC-2000; 2000US-0256188P.
 21-DEC-2000; 2000US-0257488P.
 19-JAN-2001; 2001US-0262839P.
 26-JAN-2001; 2001US-0264402P.
 XX (INCY-) INCYTE GENOMICS INC.
 PA
 XX Tang YT, Griffin JA, Yue H, Lee EA, Baughn MR, Duggan BM;
 PI Wallia NK, Lee S, Ramkumar J, Warren BA, Gandhi AR, Lu Y;
 PI Yao MG, Ding L, Tribouley CM, Sanjanwala MM, Arvizu C, Hillman JL;
 XX WPI; 2002-537565/57.
 DR P-PSDB; AAE25377.
 XX
 XX Novel human enzyme, NZMS useful in diagnosis, prevention or treatment of
 PT cell proliferative, autoimmune/inflammatory, cardiovascular, and
 PT gastrointestinal, neurological, pulmonary, reproductive and eye
 PT disorders.
 XX
 XX Claim 74; Page 159-160; 173pp; English.
 PS
 XX The invention relates to human enzymes designated NZMS and nucleic acid
 CC molecules encoding such proteins. Sequences of the invention are useful
 CC for diagnosing, treating or preventing disorders associated with aberrant
 CC expression of NZMS. The disorders treated include cell proliferative
 CC disorders such as hepatitis, psoriasis, cancer (e.g. leukaemia),
 CC autoimmune disorders such as diabetes, acquired immune deficiency
 CC syndrome (AIDS)), cardiovascular disorders such as arteriosclerosis,
 CC hypertension), gastrointestinal disorders (e.g. anorexia, gastritis),
 CC neurological disorders (e.g. epilepsy, dementia), pulmonary disorders
 CC (e.g. emphysema, asthma), reproductive or eye disorders. Polypeptides of
 CC the invention is useful in a number of drug screening techniques and to
 CC analyse the proteome of a tissue or cell type. They are also useful as
 CC elements on a microarray. Polynucleotides of the invention are useful for
 CC creating knockin humanised animals or transgenic animals to model human
 CC diseases, in somatic or germline gene therapy, to generate a transcript
 CC image of a tissue or cell type, for detecting difference in the
 CC chromosomal location due to translocation or inversion among normal,
 CC carrier or affected individuals and as hybridisation probes for mapping
 CC naturally occurring genomic sequences. The present sequence is human NZMS
 CC cDNA
 XX
 SQ Sequence 1023 BP; 287 A; 253 C; 211 G; 272 T; 0 U; 0 Other;
 Query Match 81.9%; Score 745.8; DB 6; Length 1023;
 Best Local Similarity 97.2%; Pred.No. 3.2e-232;
 Matches 759; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
 QY 17 TTCCACCCCGAGGACCATGTGAGGCTCAGCTGGGGATACCGGAGCACAACGGTCTTA 76
 DB 1 TTCCACCCCGAGGACCATGTGAGGCTCAGCTGGGGATACCGGAGCACAACGGTCTTA 60
 QY 77 TTCACCTGGAAGAAATTTTCCCTATTGCTGATGCTGATCAGCAATCTCCAATTGAGATTA 136
 DB 61 TTCACCTGGAAGAAATTTTCCCTATTGCTGATGCTGATCAGCAATCTCCAATTGAGATTA 120
 QY 137 AAACCAAGAGTGAATATGACTTCCCTCCGACCACTTAGTATCAAGTATGACCCAA 196
 DB 121 AAACCAAGAGTGAATATGACTTCCCTCCGACCACTTAGTATCAAGTATGACCCAA 180
 QY 197 GCTCAGTAAAATCATCAGCAACAGCGGCCATTCCTTCAATGTGACTTTGATGACACAG 256
 DB 181 GCTCAGTAAAATCATCAGCAACAGCGGCCATTCCTTCAATGTGACTTTGATGACACAG 240
 QY 257 AGAACAATATGATTCGTGGTGGTGCCTCTCTCACTGGAAGCTACAGTTACGGCAGGTTTC 316
 DB 241 AGAACAATATGATTCGTGGTGGTGCCTCTCTCACTGGAAGCTACAGTTACGGCAGGTTTC 300
 QY 317 ACCTTCACCTGGGGGTCGGCTGATGACCAACCGGCTCCGAGGCACATAGTAGATGAGTGAGCT 376

DB 301 ACCTTCACCTGGGGTCCGCTGATGACCAACCGCTCCGAGCACAATAGTAGATGAGTGAGCT 360
 QY 377 ATGCTGCAGAGCTCCATGTTGTTCACTGGAATTCAGACAAATACCCAGCTTTGTTGAGG 436
 DB 361 ATGCTGCAGAGCTCCATGTTGTTCACTGGAATTCAGACAAATACCCAGCTTTGTTGAGG 420
 QY 437 CAGCTCATGAACACAGATGACATGGCTGCTTGGAGTGTGTTTACAGATTGGTGAACCTA 496
 DB 421 CAGCTCATGAACACAGATGACATGGCTGCTTGGAGTGTGTTTACAGATTGGTGAACCTA 480
 QY 497 ATTCCCAACTGCACAAAGATTACTGACACATTTCCGATTTCAATTAAGAAAAAGGGTAAACAAA 556
 DB 481 ATTCCCAACTGCACAAAGATTACTGACACATTTCCGATTTCAATTAAGAAAAAGGGTAAACAAA 540
 QY 557 CTCGATTACAAAATTTTGACCTATTGCTCTCTCTCCACCATCTCGGGACTACTGGACAT 616
 DB 541 CTCGATTACAAAATTTTGACCTATTGCTCTCTCTCCACCATCTCGGGACTACTGGACAT 600
 QY 617 ATCTGGTTCCTTACAGTTCCACCTCTCTTCTGAGAGTGTCACTGATGTTTAAAGC 676
 DB 601 ATCTGGTTCCTTACAGTTCCACCTCTCTTCTGAGAGTGTCACTGATGTTTAAAGC 660
 QY 677 AACCTATAAACATCAGCTCTCAACAGCTGCGCAAAATTTCCGAGTCTCTCTGTGCACAGCG 736
 DB 661 AACCTATAAACATCAGCTCTCAACAGCTGCGCAAAATTTCCGAGTCTCTCTGTGCACAGCG 720
 QY 737 AGGGTGAAGCAGCAGCTTTTCTGTGATGAGTCTCACTCTGTCAACCGAGCTGGAGGCA 796
 DB 721 AGGGTGAAGCAGCAGCTTTTCTGTGATGAGTCTCACTCTGTCAACCGAGCTGGAGGCA 780
 QY 797 G 797
 DB 781 G 781
 RESULT 5
 APT33350
 ID APT33350 standard; DNA; 828 BP.
 XX
 AC APT33350;
 XX
 DT 22-MAY-2003 (first entry)
 XX
 DE NOVX DNA sequence SEQ ID No 15.
 XX
 KW Hepatotropic; immunosuppressive; cardiant; hypertensive; tranquilizer;
 KW vulnary; virucide; antibacterial; protozoacide; fungicide; nootropic;
 KW antiparasitic; neuroprotective; cerebroprotective; antiparkinsonian;
 KW anticonvulsant; antiaddictive; analgesic; dermatological; keratolytic;
 KW antiseborrheic; antirheumatic; antiarthritic; antiinflammatory; anti-HIV;
 KW cytostatic; antisthmatic; antipsoriatic; hypotensive; osteopathic;
 KW antiulcer; anorectic; antidiabetic; antiallergic; haemostatic;
 KW neuroleptic; antidepressant; antiinfertility; NOVX; human disease;
 KW NOVX-associated disorder; trauma; viral; bacterial; fungal; protozoal;
 KW parasitic infection; Alzheimer's disease; stroke; forensic biology;
 KW immunogen; non-human transgenic animal; gene therapy; gene; ds.
 XX
 XX Unidentified.
 XX
 XX WO200281517-A2.
 XX
 PD 17-OCT-2002.
 XX
 PD 22-JAN-2002; 2002WO-US002064.
 PF
 XX 19-JAN-2001; 2001US-0262892P.
 PR 23-JAN-2001; 2001US-0263598P.
 PR 24-JAN-2001; 2001US-0263799P.
 PR 25-JAN-2001; 2001US-0264117P.
 PR 25-JAN-2001; 2001US-0264139P.
 PR 26-JAN-2001; 2001US-0264478P.
 PR 30-JAN-2001; 2001US-0263351P.

02-MAR-2001; 2001US-0272870P.
 14-MAR-2001; 2001US-0275927P.
 14-MAR-2001; 2001US-0275990P.
 15-MAR-2001; 2001US-0276449P.
 20-MAR-2001; 2001US-0277358P.
 23-MAR-2001; 2001US-0278151P.
 29-MAR-2001; 2001US-0279857P.
 20-APR-2001; 2001US-0285140P.
 20-APR-2001; 2001US-0285141P.
 30-APR-2001; 2001US-0287484P.
 17-MAY-2001; 2001US-0291701P.
 08-JUN-2001; 2001US-0296960P.
 10-JUL-2001; 2001US-0304353P.
 10-JUL-2001; 2001US-0304355P.
 12-JUL-2001; 2001US-0304886P.
 09-AUG-2001; 2001US-0311289P.
 13-AUG-2001; 2001US-0311975P.
 16-AUG-2001; 2001US-0312937P.
 18-OCT-2001; 2001US-0330227P.
 29-NOV-2001; 2001US-0334198P.
 (CURA-) CURAGEN CORP.

Decristofaro MF, Padigaru M, Miller C, Tchernev V, Zhong H;
 Zhong M, Anderson D, Ballinger R, Gerlach V, Spytek KA, Rastelli L;
 Kekuda R, Guo X, Zehrhusen B, Andrew D, Mezes P, Patturajan M;
 Burgess CE, Eisen A, Wolenc A, Baumgartner J, Shimkets RA, Gusev V;
 Vernet CAM, Taupier RJ, Pena C, Shenoy S, Li L, Casman S, Boldog F;
 Fernandes E, Smithson G, Malyankar U, Taillon B, Liu X;

WPI; 2003-058504/05.

P-PSDB; ABJ37885.

New polypeptides, designated as NOVX, useful for diagnosing and treating
 infections, neurological diseases, cancer, allergy, and bone,
 immunological, skin, renal, brain, muscle and autoimmune disorders.

Claim 9; Page 65-66; 672pp; English.

The invention relates to a novel isolated polypeptide, designated NOVX
 (NOV1 - 33), consisting of a mature form of one of 61 sequences, given in
 the specification, or its variant, where amino acid residue(s) in the
 variant differ from the mature form, provided that the variant differs in
 not more than 15 % of the amino acids from the sequence of the mature
 form. The NOVX polypeptides, nucleic acids encoding the polypeptides, and
 an antibody to the polypeptides, are useful for treating or preventing a
 NOVX-associated disorder in humans and for treating a syndrome associated
 with a human disease (NOVX-associated disorder). NOVX polypeptides and
 the encoding nucleic acids, are useful for determining the presence of or
 predisposition to a disease associated with altered levels of NOVX
 polypeptide and polynucleotide, by measuring the level of polypeptide
 expression or the amount of nucleic acid from a mammal and comparing it
 with another mammal not having or not predisposed to the disease. NOVX
 polypeptide is also useful for identifying an agent that binds to NOVX
 and a cell expressing NOVX is useful for identifying an agent that
 modulates the expression or activity of NOVX. The antibodies and a
 polypeptide having 95 % sequence identity to NOVX polypeptide are useful
 for treating a pathological state in a mammal. The antibodies are also
 useful for determining the presence or amount of NOVX in a sample. NOVX
 polypeptides, polynucleotides and antibodies specific for the
 polypeptides are useful for treating or preventing disorders or syndromes
 including trauma, viral, bacterial, fungal, protozoal, and parasitic
 infections. They can also treat disorders such as e.g., Alzheimer's
 disease or a stroke. The NOVX encoding nucleic acids are useful for
 expressing the NOVX proteins, to detect NOVX mRNA, or a genetic lesion in
 a NOVX gene and to modulate NOVX activity. NOVX sequences are also useful
 for identifying a cell or tissue type in a biological sample, to amplify
 DNA sequences from very small biological samples such as tissues e.g.
 hair or skin or body fluids in forensic biology and as primers and probes
 for use in identifying and/or cloning NOVX homologues in other cell
 types. The NOVX proteins are useful as an immunogen to generate
 antibodies which are useful for diagnostically monitoring protein levels
 and modulating NOVX activity. Cells comprising NOVX nucleic acids are

CC useful for producing non-human transgenic animals which are useful for
 CC studying the function and/or activity of NOVX protein and for identifying
 CC and/or evaluating modulators of NOVX protein activity. The NOVX nucleic
 CC acids can be used in gene therapy. This polynucleotide sequence
 CC represents a NOVX DNA sequence of the invention

XX SQ Sequence 828 BP; 224 A; 214 C; 181 G; 209 T; 0 U; 0 Other;

Query Match 81.3%; Score 740.6; DB 7; Length 828;

Best Local Similarity 96.9%; Pred. No. 1.4e-230; Indels 0; Gaps 0;

Matches 755; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 19 CCACCCCGAGGACCATGTCGAGGCTCAGCTGGGATACCGGAGCACACGTCCTATT 78

Db 1 CCACCCCGAGGACCATGTCGAGGCTCAGCTGGGATACCGGAGCACACGTCCTATT 60

QY 79 CACTGGAAGGAATTTTCCCTATTGCTGATGTGATCAGCAATCTCCAATTGAGATTAA 138

Db 61 CACTGGAAGGAATTTTCCCTATTGCTGATGTGATCAGCAATCTCCAATTGAGATTAA 120

QY 139 ACCAAGAGTGAAATATGACTCTCCCTCCGACCACTTAGTATCAAGTATGACCCAAGC 198

Db 121 ACCAAGAGTGAAATATGACTCTCCCTCCGACCACTTAGTATCAAGTATGACCCAAGC 180

QY 199 TCAGCTTAAATCATCAGCAACAGCGGCCATTCTTCAATGTTGACTTTGATGACACAGAG 258

Db 181 TCAGCTTAAATCATCAGCAACAGCGGCCATTCTTCAATGTTGACTTTGATGACACAGAG 240

QY 259 AACAAATCAGTTCTCGTGGTGGTCTCTCCTGGAAGCTACAGGTTAGCGGAGTTTAC 318

Db 241 AACAAATCAGTTCTCGTGGTGGTCTCTCCTGGAAGCTACAGGTTAGCGGAGTTTAC 300

QY 319 CTTCACTGGGGTCCGCTGATGACACACGCTCCGAGCACATAGTAGAGTGAGCTAT 378

Db 301 CTTCACTGGGGTCCGCTGATGACACACGCTCCGAGCACATAGTAGAGTGAGCTAT 360

QY 379 GCTGCAGAGCTCCATGTTGTTTCACTGGAATTCAGACAAATACCCAGCTTTGTTGAGGCA 438

Db 361 GCTGCAGAGCTCCATGTTGTTTCACTGGAATTCAGACAAATACCCAGCTTTGTTGAGGCA 420

QY 439 GCTCATGAACCGATGGACTGGCTGTTGGGAGTGTCTTTTACAGATTGGTGAACCTAAT 498

Db 421 GCTCATGAACCGATGGACTGGCTGTTGGGAGTGTCTTTTACAGATTGGTGAACCTAAT 480

QY 499 TCCCACTCCAAAGATTACTGACACTTTGGATTCCATTAAAGAAAAGGTTAAACAACT 558

Db 481 TCCCACTCCAAAGATTACTGACACTTTGGATTCCATTAAAGAAAAGGTTAAACAACT 540

QY 559 CGATTACAAATTTTGAACCTATTGCTCTGCTTCCACCATCTCTGGGACTACTGGACATAT 618

Db 541 CGATTACAAATTTTGAACCTATTGCTCTGCTTCCACCATCTCTGGGACTACTGGACATAT 600

QY 619 CTTGGTTCCTTACAGTTTCACTCTTCTTGAGAGTGTACATGGATTGTTTAAAGCAA 678

Db 601 CTTGGTTCCTTACAGTTTCACTCTTCTTGAGAGTGTACATGGATTGTTTAAAGCAA 660

QY 679 CCTATAACATCAGCTCTCAACAGCTGGCCAAATTTTCGAGTCTCTGTGCACAGCGGAG 738

Db 661 CCTATAACATCAGCTCTCAACAGCTGGCCAAATTTTCGAGTCTCTGTGCACAGCGGAG 720

QY 739 GGTGAAGCAGCAGCTTTTCTGTGATAGTCTCACTCTGTCAACCCAGGTGGAGGGCAG 797

Db 721 GGTGAAGCAGCAGCTTTTCTGTGATAGTCTCACTCTGTCAACCCAGGTGGAGGGCAG 779

RESULT 6

ABT33351

ID ABT33351 standard; DNA; 833 BP.

XX AC ABT33351;

XX AC ABT33351;

XX 22-MAY-2003 (first entry)

XX

NOVX DNA sequence SEQ ID No 17.

DE Hepatotropic; immunosuppressive; cardiant; hypertensive; tranquilizer;
 XX vulnary; viricide; antibacterial; protozoacide; fungicide; nootropic;
 KW antiparasitic; neuroprotective; cerebroprotective; antiparkinsonian;
 KW anticonvulsant; antidiabetic; analgesic; dermatological; keratolytic;
 KW antieborrheic; antihemetic; antiarthritic; antiinflammatory; anti-HIV;
 KW cytostatic; antiasthmatic; antipsoriatic; hypotensive; osteopathic;
 KW antiulcer; anorectic; antidiabetic; antiallergic; haemostatic;
 KW neuroleptic; antidepressant; antinfertility; NOVX; human disease;
 KW NOVX-associated disorder; trauma; viral; bacterial; fungal; protozoal;
 KW parasitic infection; Alzheimer's disease; stroke; forensic biology;
 KW immunogen; non-human transgenic animal; gene therapy; gene; ds.
 XX Unidentified.

OS WO200281517-A2.
 XX 17-OCT-2002.

PN 22-JAN-2002; 2002WO-US002064.
 XX 19-JAN-2001; 2001US-0262892P.
 PR 23-JAN-2001; 2001US-0263598P.
 PR 24-JAN-2001; 2001US-0263799P.
 PR 25-JAN-2001; 2001US-0264117P.
 PR 26-JAN-2001; 2001US-0264139P.
 PR 26-JAN-2001; 2001US-0264478P.
 PR 30-JAN-2001; 2001US-0263351P.
 PR 02-MAR-2001; 2001US-0272870P.
 PR 14-MAR-2001; 2001US-0275927P.
 PR 14-MAR-2001; 2001US-0275990P.
 PR 15-MAR-2001; 2001US-0278449P.
 PR 20-MAR-2001; 2001US-0277358P.
 PR 23-MAR-2001; 2001US-0278151P.
 PR 29-MAR-2001; 2001US-0279857P.
 PR 20-APR-2001; 2001US-0285140P.
 PR 30-APR-2001; 2001US-0285141P.
 PR 20-APR-2001; 2001US-0287484P.
 PR 17-MAY-2001; 2001US-0291701P.
 PR 08-JUN-2001; 2001US-0296960P.
 PR 10-JUL-2001; 2001US-0304353P.
 PR 12-JUL-2001; 2001US-0304355P.
 PR 12-JUL-2001; 2001US-0304886P.
 PR 09-AUG-2001; 2001US-0311289P.
 PR 13-AUG-2001; 2001US-0311975P.
 PR 16-AUG-2001; 2001US-0312937P.
 PR 18-OCT-2001; 2001US-0330227P.
 PR 29-NOV-2001; 2001US-0334198P.

(CURA-) CURAGEN CORP.

PI Decristofaro MF, Padigaru M, Miller C, Tchernev V, Zhong H;
 PI Zhong M, Anderson D, Ballinger R, Gerlach V, Spytek XA, Rastelli L;
 PI Kekuda R, Guo X, Zerkhus B, Andrew D, Mezes P, Patturajan M;
 PI Burgess CE, Eisen A, Wolenc A, Baumgartner J, Shinkets RA, Gusev V;
 PI Vernet CAM, Taupier RJ, Pena C, Shenoy S, Li L, Casman S, Boldog F;
 PI Fernandes E, Smithson G, Malyankar U, Tallon B, Liu X;
 XX WPI; 2003-058504/05.
 DR P-PSDB; ABJ37886.

XX New polypeptides, designated as NOVX, useful for diagnosing and treating
 PT infections, neurological diseases, cancer, allergy, and bone,
 PT immunological, skin, renal, brain, muscle and autoimmune disorders.
 XX Claim 9; Page 66-67; 672pp; English.

PS The invention relates to a novel isolated polypeptide, designated NOVX
 CC (NOVI - 33), consisting of a mature form of one of 61 sequences, given
 CC the specification, or its variant, where amino acid residue(s) in the
 CC variant differ from the mature form, provided that the variant differs in
 CC not more than 15 % of the amino acids from the sequence of the mature

form. The NOVX polypeptides, nucleic acids encoding the polypeptides, and
 an antibody to the polypeptides, are useful for treating or preventing a
 NOVX-associated disorder in humans and for treating a syndrome associated
 with a human disease (NOVX-associated disorder). NOVX polypeptides and
 the encoding nucleic acids, are useful for determining the presence of or
 predisposition to a disease associated with altered levels of NOVX
 polypeptide and polynucleotide, by measuring the level of polypeptide
 expression or the amount of nucleic acid from a mammal and comparing it
 with another mammal not having or not predisposed to the disease. NOVX
 polypeptide is also useful for identifying an agent that binds to NOVX
 and a cell expressing NOVX is useful for identifying an agent that
 modulates the expression or activity of NOVX. The antibodies and a
 polypeptide having 95 % sequence identity to NOVX polypeptide are useful
 for treating a pathological state in a mammal. The antibodies are also
 useful for determining the presence or amount of NOVX in a sample. NOVX
 polypeptides, polynucleotides and antibodies specific for the
 polypeptides are useful for treating or preventing disorders or syndromes
 including trauma, viral, bacterial, fungal, protozoal, and parasitic
 infections. They can also treat disorders such as e.g., Alzheimer's
 disease or a stroke. The NOVX encoding nucleic acids are useful for
 expressing the NOVX proteins, to detect NOVX mRNA, or a genetic lesion in
 a NOVX gene and to modulate NOVX activity. NOVX sequences are also useful
 for identifying a cell or tissue type in a biological sample, to amplify
 DNA sequences from very small biological samples such as tissues e.g.
 hair or skin or body fluids in forensic biology and as primers and probes
 for use in identifying and/or cloning NOVX homologues in other cell
 types. The NOVX proteins are useful as an immunogen to generate
 antibodies which are useful for diagnostically monitoring protein levels
 and modulating NOVX activity. Cells comprising NOVX nucleic acids are
 useful for producing non-human transgenic animals which are useful for
 studying the function and/or activity of NOVX protein and for identifying
 and/or evaluating modulators of NOVX protein activity. The NOVX nucleic
 acids can be used in gene therapy. This polynucleotide sequence
 CC represents a NOVX DNA sequence of the invention
 XX

Sequence 833 BP; 225 A; 210 C; 184 G; 214 T; 0 U; 0 Other;
 Query Match 80.0%; Score 728.8; DB 7; Length 833;
 Best Local Similarity 97.1%; Pred. No. 1e-226;
 Matches 742; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 34 ATGTCGAGGCTCAGTGGGGATACCGGAGGACACACGGTCTCTATTCTCTGGGAGGATTT 93
 DB 21 ATGTCGAGGCTCAGTGGGGATACCGGAGGACACACGGTCTCTATTCTCTGGGAGGATTT 80
 QY 94 TTCCCTATTGCTGATGGTGATCAGCAATCTCCAAATTGAGATTAAACCAAGAGTGA 153
 DB 81 TTCCCTATTGCTGATGGTGATCAGCAATCTCCAAATTGAGATTAAACCAAGAGTGA 140
 QY 154 TATGACTCTTCCCTCCGACCTTAGTATCAAGTATGACCCAGCTCAGCTAAATCATC 213
 DB 141 TATGACTCTTCCCTCCGACCTTAGTATCAAGTATGACCCAGCTCAGCTAAATCATC 200
 QY 214 AGCAACAGGCGCATTCCTTCAATGTTGATGATGACAGAGAACAAATCATCTCTG 273
 DB 201 AGCAACAGGCGCATTCCTTCAATGTTGATGATGACAGAGAACAAATCATCTCTG 260
 QY 274 CGTGTGCTCTCTCCTCCTGGAAGCTTACAGGTTACGGCAGGTTACCTTCTCTGGGGTCC 333
 DB 261 CGTGTGCTCTCTCCTCCTGGAAGCTTACAGGTTACGGCAGGTTACCTTCTCTGGGGTCC 320
 QY 334 GCTGATGACACGGCTCCGAGCAGATAGTAGATGAGGTGAGTATGCTGCGAGAGTCCAT 393
 DB 321 GCTGATGACACGGCTCCGAGCAGATAGTAGATGAGGTGAGTATGCTGCGAGAGTCCAT 380
 QY 394 GTTGTTCACCTGGAATTCAGACAAATACCCAGCTTTGTTGAGGAGCTCATGAACAGAT 453
 DB 381 GTTGTTCACCTGGAATTCAGACAAATACCCAGCTTTGTTGAGGAGCTCATGAACAGAT 440
 QY 454 GGACTGGCTGTCTTGGGAGTGTGTTTACAGATTGGTGAACCTAATTCCTCAACTGCAAAAG 513
 DB 441 GGACTGGCTGTCTTGGGAGTGTGTTTACAGATTGGTGAACCTAATTCCTCAACTGCAAAAG 500

514 ATTACTGACATTGGATTCCATTAAAGAAAGGGTAAACAACTCGATTACAAATTTT 573
501 ATTACTGACACTTGGATTCCATTAAAGAAAGGGTAAACAACTCGATTACAAATTTT 560
574 GACCTATTGCTCTGCTCCACCATCTGGGACTACTGGACATATCTCTGTTCTTACA 633
561 GACCTATTGCTCTGCTCCACCATCTGGGACTACTGGACATATCTCTGTTCTTACA 620
634 GTTCACCTCTCTTCTGAGAGTGTACATGGATTGTTTTAAAGCAACCTATAACATCAGC 693
621 GTTCACCTCTCTTCTGAGAGTGTACATGGATTGTTTTAAAGCAACCTATAACATCAGC 680
694 TCTCAACAGCTGGCCAAATTTCCGAGTCTCTCTGTGTCACAGCGAGGTGAACAGAGCT 753
681 TCTCAACAGCTGGCCAAATTTCCGAGTCTCTCTGTGTCACAGCGAGGTGAACAGAGCT 740
754 TTTCTGTGTAGAGTCTCACTGTGTCACCCAGGCTGGAGGGCAG 797
741 TTTCTGTGTAGAGTCTCACTGTGTCACCCAGGCTGGAGGGCAG 784

RESULT 7
AAF22342
ID AAF22342 standard; cDNA; 688 BP.
AC AAF22342;
XX
XX 26-MAR-2001 (first entry)
XX Human secreted protein gene 27 SEQ ID NO:37.
XX
XX Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
XX antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
XX cerebroprotective; neutrophic; neuroprotective; antibacterial; virucide;
XX fungicide; ophthalmological; vulnary; gene therapy; neoplasm;
XX autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;
XX cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
XX cerebral ischaemia; angiogenesis; nervous system disorder; infection;
XX Alzheimer's disease; ocular disorder; corneal infection; wound healing;
XX skin aging; food additive; preservative; ss.
XX
OS Homo sapiens.
XX
XX WO20061748-A1.
XX
XX 19-OCT-2000.
XX
XX 06-APR-2000; 2000WO-US008982.
XX
XX 09-APR-1999; 99US-0128696P.
XX 14-JAN-2000; 2000US-0176069P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM, Komatsoulis G;
XX
XX WPI; 2000-638566/61.
XX P-PSDB; AAB63075.
XX
XX New nucleic acid molecules encoding 48 human secreted proteins for
XX diagnosing, preventing, treating or ameliorating medical conditions and
XX used as food additives or preservatives.
XX
XX Claim 1; Page 417; 480pp; English.
XX
XX AAF22316 to AAF22363 encode the human secreted proteins given in AAB63049
XX to AAB63096. AAB63097 to AAB63132 represent more human secreted proteins
XX and polypeptides homologous to them. Human secreted proteins have
XX activities based on the tissues and cells the genes are expressed in.
XX Examples of activities include: immunosuppressive; antiarthritic;
XX antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
XX cerebroprotective; neutrophic; neuroprotective; antibacterial; virucide;
XX fungicide; ophthalmological; and vulnary. The polynucleotides and

proteins can be used to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used in diagnosing a pathological condition or susceptibility to a pathological condition. Disorders which are diagnosed or treated include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g. Alzheimer's disease, infections caused by bacteria, viruses and fungi and ocular disorders e.g. corneal infection. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, in supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities, fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors and other nutritional components. AAF22307 to AAF22315 and AAB63048 represent sequences used in the exemplification of the present invention

XX
XX
SQ Sequence 688 BP; 192 A; 174 C; 150 G; 172 T; 0 U; 0 Other;

Query Match 54.1%; Score 493.2; DB 3; Length 688;
Best Local Similarity 95.7%; Pred. No. 6.9e-150;
Matches 507; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 268 GTTCTGCGTGGTGGTCTCTCACTGGAAGCTACAGGTTACGGCAGGTTCCACCTTCACTGG 327
DB 43 GTTCTGCGTGGTGGTCTCTCACTGGAAGCTACAGGTTACGGCAGGTTCCACCTTCACTGG 102

QY 328 GGGTCCGCTGATGACCCAGGCTCCGAGCACATAGTAGAGTGGAGTATGTCGACAG 387
DB 103 GGGTCCGCTGATGACCCAGGCTCCGAGCACATAGTAGAGTGGAGTATGTCGACAG 162

QY 388 CTCATGTTTCTACCTGGAATTCAGACAAATCCCGAGCTTGTTCAGGAGCTCATGAA 447
DB 163 CTCATGTTTCTACCTGGAATTCAGACAAATCCCGAGCTTGTTCAGGAGCTCATGAA 222

QY 448 CCAGATGAGTCTGCTCTCTGGAGTGTGTTTACAGATTGGTGAACTAATTCCTCAACTG 507
DB 223 CCAGATGAGTCTGCTCTCTGGAGTGTGTTTACAGATTGGTGAACTAATTCCTCAACTG 282

QY 508 CAAAGATTACTGACACTTTGGATTCCATTAAAGAAAAGGGTAAACAACTCGATTACACA 567
DB 283 CAAAGATTACTGACACTTTGGATTCCATTAAAGAAAAGGGTAAACAACTCGATTACACA 342

QY 568 AATTTGACCTATTGCTCTGTTCCACCATCTGGAGTCTCTGGAGTCTCTGGAGTCTCT 627
DB 343 AATTTGACCTATTGCTCTGTTCCACCATCTGGAGTCTCTGGAGTCTCTGGAGTCTCT 402

QY 628 CTTTACAGTTCACCTCTCTTGGAGTGTGTCACATGATTGTTTTAAAGCAACCTATAAAC 687
DB 403 CTTTACAGTTCACCTCTCTTGGAGTGTGTCACATGATTGTTTTAAAGCAACCTATAAAC 462

QY 688 ATCAGCTCTCAACAGCTGGCCAAATTTGCGAGTCTCTCTGTGTCACAGGGAGGTGAAGCA 747
DB 463 ATCAGCTCTCAACAGCTGGCCAAATTTGCGAGTCTCTCTGTGTCACAGGGAGGTGAAGCA 522

QY 748 GCAGCTTTTCTGTGTAGAGTCTCTCACTCTGTACCCAGGCTGGAGGGCAG 797
DB 523 GCAGCTTTTCTGTGTAGAGTCTCTCACTCTGTACCCAGGCTGGAGGGCAG 572

RESULT 8
AAB68431/c
ID AAB68431 standard; cDNA; 462 BP.
XX
XX AAB68431;
XX
XX 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #4235.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

food supplement; medical imaging; diagnostic; genetic disorder; ss.
Homo sapiens.
WO200175067-A2.
11-OCT-2001.
30-MAR-2001; 2001WO-US008631.
31-MAR-2000; 2000US-00540217.
23-AUG-2000; 2000US-00649167.
(HYSE-) HYSBQ INC.
Drmanac RT, Liu C, Tang YT;
WPI; 2001-639362/73.
P-PSDB; ABG04244.
New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity.
Claim 1; SEQ ID NO 4235; 103pp; English.
The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention. Note: the sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
Sequence 462 BP; 113 A; 115 C; 113 G; 121 T; 0 U; 0 Other;
Query Match 48.8%; Score 444.4; DB 5; Length 462;
Best Local Similarity 99.8%; Pred. No. 4.6e-134;
Matches 445; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 23 CCCGAGGACCATGTCGAGGCTCAGTGGGATACCGGAGCACACGCTCTATTCACT 82
DB 446 CCCGAGGACCATGTCGAGGCTCAGTGGGATACCGGAGCACACGCTCTATTCACT 387
QY 83 GGAAGGAATTTTCCCTATTGCTGATGGTGTATCAGCAATCTCCAATTGAGATTAAACCA 142
DB 386 GGAAGGAATTTTCCCTATTGCTGATGGTGTATCAGCAATCTCCAATTGAGATTAAACCA 327
QY 143 AAGAAGTGAATATGACTCTCCCTCCGACCACTTAGTATCAAGTATGACCAAGCTCAG 202
DB 326 AAGAAGTGAATATGACTCTCCCTCCGACCACTTAGTATCAAGTATGACCAAGCTCAG 267
QY 203 CTAAATCATCAGCACAGCGGCAATCTCTCAATGTTGACTTTCATGACACAGAGACA 262
DB 266 CTAAATCATCAGCACAGCGGCAATCTCTCAATGTTGACTTTCATGACACAGAGACA 207
QY 263 AATCATGTTCTGGTGGTGGTCTCTCTCACTGGAGGCTACAGGTTACGGAGGTTCACTTC 322

Db 206 AATCAGTTCTGCGGGTGGTCTCTCACTGAAAGTACAGGTTACGGCAGGTTCACTTC 147
QY 323 ACTGGGGTCCGCTGATGACACCGGCTCCGAGCACATAGTAGTAGTGGAGTGGCTATGCTG 382
Db 146 ACTGGGGTCCGCTGATGACACCGGCTCCGAGCACATAGTAGTAGTGGAGTGGCTATGCTG 87
QY 383 CAGAGCTCCATGTTGTTTCTACTGGAATTCAGACAAATACCCAGCTTTGTTGAGGAGCTC 442
Db 86 CAGAGCTCCATGTTGTTTCTACTGGAATTCAGACAAATACCCAGCTTTGTTGAGGAGCTC 27
QY 443 ATGAACACAGATGACGCTGGCTGCTTG 468
Db 26 ATGAACACAGATGACGCTGGCTGCTTG 1
RESULT 9
ACH26864
ID ACH26864 standard; cDNA; 496 BP.
XX AC ACH26864;
XX DT 13-OCT-2003 (first entry)
XX DE Human adult ovary cDNA #5244.
XX KW Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
XX KW genome mapping; biodiversity; genetic disorder.
XX OS Homo sapiens.
XX PN US2003073623-A1.
XX PD 17-APR-2003.
XX PF 30-JUL-2001; 2001US-00918995.
XX PR 30-JUL-2001; 2001US-00918995.
XX PA (DRMA/) DRMANAC R T.
XX PA (LABA/) LABAT I.
XX PA (STAC/) STACHE-CRAIN B.
XX PA (DICK/) DICKSON M C.
XX PA (JONE/) JONES L W.
XX PI Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX DR WPI; 2003-615964/58.
XX PT New polynucleotide sequences obtained from various cDNA libraries, useful as hybridization probes, as oligomers for PCR, for chromosome and gene mapping, in the recombinant production of protein, or in generating antisense DNA or RNA.
PS Claim 1; SEQ ID NO 14076; 44pp; English.
XX The invention relates to an isolated polynucleotide comprising any one of 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was determined by the technique of SBH (sequencing by hybridisation). Also included is a purified polypeptide comprising a sequence corresponding to a reading frame of the novel polynucleotide. The nucleic acid sequences are useful in diagnostics as expressed sequence tags (EST) for identifying expressed genes or for physical mapping of the human genome, in forensics, in assessing biodiversity, or in identifying mutations responsible for genetic disorders and other traits. The nucleotide sequences are also useful as hybridisation probes, as oligomers for PCR, for chromosome and gene mapping, in the recombinant production of protein, or in generating antisense DNA or RNA. The purified polypeptide is useful for generating antibodies specific for it. The present sequence is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from USPTO at seqdata.uspto.gov/sequence.html?DocID=20030073623

isolated nucleic acids and polypeptides, useful for preventing diagnosing

Db 484 TCAGTAGCTCACTGGAATCTGCAAGTACTCCAGCCTTCTGAGCTGCCTCAAAGGC 543
 QY 450 AGATGGACTGGCTGCTCTGGGAGTGTGTTTACAGATTGGTGAACCTTAATTCCTCAACTGCA 509
 Db 544 TGATGGTTGGCAGTATTGGTGTGTTGATGAAGGTTGGTGAAGCCCAACCAAGCTGCA 603
 QY 510 AAAGATTACTGACACTTTGGATTCCATTAAGAAAGGGTAAACAACTCGATTACAAA 569
 Db 604 GAAAGTACTTGTATGGCTTCAAGAGCAATTAACCAAGGGCAACGAGGCCATTACAAA 663
 QY 570 TTTTGACCTATTGTTCTGCTTCCACCATCCCTGGGACTACTGACATATCCCTGGTCTCT 629
 Db 664 TTTTGACCCCTACT 723
 QY 630 TACAGTTCACCT 689
 Db 724 GACTCATCT 783
 QY 690 CAGCTCTCAAGCTGGCCAAATTTGGCAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 749
 Db 784 CAGCTCAGAGCAGCTGGCAAAATTCGCGAGCCTTCTATCAAAATGTTGAAGGTGATAACGC 843
 QY 750 AGCTTTTCTGTATAGAGTCTCACTCTGTCACCCAGGCTGGAGGGCAG 797
 Db 844 TGTCCTCATGAGCAGCAACAACCGGCCCAACCACTCTGAAGGGCAG 891

RESULT 14

AAD59160

ID AAD59160 standard; cDNA; 2826 BP.

XX AAD59160;

AC AAD59160;

DT 18-DEC-2003 (first entry)

DE Human carbonic anhydrase I (CAI) cDNA.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Matches 494; Conservative 0; Mismatches 274; Indels 0; Gaps 0;
 QY 30 GACCATGTGCGAGCTCAGCTGGGATACCGGAGCACAACCGTCTCTATTCACTTGAAGCA 89
 Db 1218 GATAATGCAAGTCCAGACTGGGATATGATGACAAAATGGTCTGAAACATGGAGCAA 1277
 QY 90 ATTTTCCCTATTGCTGATGGTGTGATCAGCAATCTCAATTTGAGATTAAACCAAGAAGT 149
 Db 1278 GCTGTATCCCATGTCGCAATTAACAGTCCCTGTTGATATTAACCAAGTGAAC 1337
 QY 150 GAAATATGACTCTTCCCTCCGACCACTTAGTATCAAGTATGACCAAGCTCAGCTAAAT 209
 Db 1338 CAAACATGACCTCTCTGAAACCTATTAGTGTCTCTTACAAACCCGACAGCCAAAGA 1397
 QY 210 CATCAGCAACGCGCCCATTCCTTCAATTTGACTTTGATGACACAGAGAAACAAATCAGT 269
 Db 1398 AATTATCAATGCGGCAATTCCTTCCATGTAATTTTGAGGACAACGATAACCGATCAGT 1457
 QY 270 TCTGCGTGGTGTCTCTCTCACTGGGAAGCTACAGGTTACGGCAGGTTCACTTCACTGGG 329
 Db 1458 GCTGAAAGGTTGGTCTCTCTCTGACAGCTACAGGCTCTTTCAGTTCCATTTTCACTGGG 1517
 QY 330 GTCGCTCATGACCAACGCTCCGAGCACATAGTAGAGTGAGTGTGAGTGTGAGTGTGAGTGT 389
 Db 1518 CAGTACAAATGAGCATGGTTAGAACATACAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 1577
 QY 390 CCAATGTTTCTCACTGGAATTCAGACAAATACCCAGCTTTTGTGAGGAGCTCATGAACC 449
 Db 1578 TCAGTAGCTCACTGGAATTCGCAAGTACTTCCAGCTTGTCTGAGGCTGCTCAAGGC 1637
 QY 450 AGATGAGCTGCTGCTTGGGAGTGTGTTTACAGATTGTTGAGTGTGAGTGTGAGTGTGAGTGT 509
 Db 1638 TGATGGTTTGGCAGTTATTGGTGTGATGAAGGTTGGTGAAGGCTTGAAGGCTGCA 1697
 QY 510 AAAGATTACTGACACTTTGGATTCCATTAAGAAAGGTTAAACAACTCGATTTCACAA 569
 Db 1698 GAAAGTACTTGAAGCTTCCAGCAATTAACCAAGGGCAACAGGCCCATTCACAAA 1757
 QY 570 TTTTGACCTATTGCTCTGCTTCCACCATCTCTGGGACTACTGGACATATCTCTGGTCTCT 629
 Db 1758 TTTTGACCCCTTACT 1817
 QY 630 TACAGTTCACCTCTTCTGAGAGTGTACATGGATTGTTTAAAGCAACCTATATAACAT 689
 Db 1818 GACTCATCTCTCTCTTATGAGAGTGTAACTTGGATCATCTGTAAGGAGCATCAGTGT 1877
 QY 690 CAGCTCTCAAGCTGGCCAAATTTTCGAGTCTCTGTCAGCGAGGGGTGAAGCAGC 749
 Db 1878 CAGCTCAGAGCAGCTGGCACAATTCGCGAGCTTCTATCAATGTTGAAGGTGATAACGC 1937
 QY 750 AGCTTTTCTGTATAGAGTCTCACTCTCTGTCACCCAGGCTGGAGGGCAG 797
 Db 1938 TGTCCTCATGAGCAGCAACAACCGGCCCAACCACTCTGAAGGGCAG 1985

RESULT 15

ABK63768

ID ABK63768 standard; cDNA; 1459 BP.

XX ABK63768;

AC ABK63768;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Novel combination of cDNAs which are differentially expressed in colon cancer, useful for detecting differential expression of one or more cDNAs in a sample containing nucleic acid samples.

Claim 1; Page 64-65; 88pp; English.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

The present invention relates to combination of cDNAs which are differentially expressed in colon cancer. The invention is useful for producing and purifying antibody, utilized as markers for treatment efficacy against colon cancer. The invention is also useful for gene therapy. The present sequence is human carbonic anhydrase I (CAI) cDNA

Sequence 2826 BP; 846 A; 633 C; 496 G; 851 T; 0 U; 0 Other;

Query Match 36.2%; Score 329.6; DB 9; Length 2826;

Best Local Similarity 64.3%; Pred. No. 46-96;

Rat sequence differentially expressed in response to a hepatotoxin #1675.

Rat; ss; hepatotoxin; expressed sequence tag; EST; drug screening;

differential expression; centrilobular necrosis; steatosis.

Rattus norvegicus.

WO200210453-A2.

07-FEB-2002.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 15:06:59 ; Search time 40 Seconds
(without alignments)
581.959 Million cell updates/sec

Title: US-10-069-434-1

Perfect score: 1289

Sequence: 1 MSRLSWGVEHNGPIHWKEF.....QLAKFRSLCTAGGAAAPL 242

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_78.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	804	62.4	261	1 CRHU1	carbonate dehydrat
2	804	62.4	261	2 JN0836	carbonate dehydrat
3	800	62.1	260	1 CRHO1D	carbonate dehydrat
4	800	62.1	261	2 JN0835	carbonate dehydrat
5	788.5	61.2	259	1 CRSH2	carbonate dehydrat
6	788.5	61.2	260	1 CRHU2	carbonate dehydrat
7	785.5	60.9	259	1 CRBO2	carbonate dehydrat
8	783.5	60.8	260	2 T08463	carbonate dehydrat
9	780	60.5	260	1 CRMQ1R	carbonate dehydrat
10	778.5	60.4	260	2 JC2580	carbonate dehydrat
11	761.5	59.1	259	1 CRRB2	carbonate dehydrat
12	754.5	58.5	259	2 A22612	carbonate dehydrat
13	754	58.5	261	2 A26344	carbonate dehydrat
14	749.5	58.1	260	1 CRMS2	carbonate dehydrat
15	748.5	58.1	260	2 JH0527	carbonate dehydrat
16	748.5	58.1	260	2 A43641	carbonate dehydrat
17	747.5	58.0	260	1 CRHU3	carbonate dehydrat
18	747.5	58.0	260	2 I52551	carbonate dehydrat
19	717.5	55.7	235	2 A22962	carbonate dehydrat
20	689	53.5	264	1 CRHU7	carbonate dehydrat
21	589	45.7	305	1 I59261	carbonate dehydrat
22	581.5	45.1	304	2 SI2579	carbonate dehydrat
23	549.5	42.6	298	2 A60519	carbonate dehydrat
24	457	35.5	200	2 A60519	carbonate dehydrat
25	453	35.1	303	2 SI2867	carbonate dehydrat
26	451	35.0	290	2 JN0576	carbonate dehydrat
27	392	30.4	304	1 CRVZ7P	cell surface-bind
28	386	29.9	304	1 CRVZW	cell surface-bind
29	386	29.9	304	2 H72162	F8L protein - vari

ALIGNMENTS

RESULT 1

CRHU1

carbonate dehydratase (EC 4.2.1.1) I [validated] - human

N:Alternate names: carbonic anhydrase I

C:Species: Homo sapiens (man)

C>Date: 07-May-1981 #sequence revision 05-May-1995 #text change 08-Dec-2000

C:Accession: JQ0786; A26573; A90668; A90180; A92128; B92147; A01138

R:Low, N.; Brady, H.J.M.; Barlow, J.H.; Sowden, J.C.; Edwards, M.; Butterworth, P.H.W.

Gene 93, 277-283, 1990

A:Title: Structure and methylation patterns of the gene encoding human carbonic anhydrases

A:Reference number: JQ0786; MUID:91033039; PMID:2121614

A:Accession: JQ0786

A:Molecule type: DNA

A:Residues: 1-261 <LOW>

A:Cross-references: GB:M33987; NID:gl79792; PIDN:AAAS1910.1; PID:gl79793

A:Experimental source: erythrocyte

R:Barlow, J.H.; Lowe, N.; Edwards, Y.H.; Butterworth, P.H.W.

Nucleic Acids Res. 15, 2386, 1987

A:Title: Human carbonic anhydrase I cDNA.

A:Reference number: A26573; MUID:87174760; PMID:3104879

A:Accession: A26573

A:Molecule type: mRNA

A:Residues: 1-261 <BAR>

A:Cross-references: GB:X05014; NID:g29599; PIDN:CAA28663.1; PID:g29600

A>Note: the authors translated the codon GAG for residue 118 as Gly

R:Gitaud, N.; Marriq, C.; Laurent-Tabusse, G.

Biochimie 56, 1031-1043, 1974

A:Title: Structure primaire de l'anhydrase carbonique erythrocytaire B humaine. III. Seq

A:Reference number: A90668; MUID:75091068; PMID:4217196

A:Accession: A90668

A:Molecule type: protein

A:Residues: 2-74, 'ND', '77-261 <GIR>

R:Andersson, B.; Nyman, P.O.; Strid, L.

Biochem. Biophys. Res. Commun. 48, 670-677, 1972

A:Title: Amino acid sequence of human erythrocyte carbonic anhydrase B.

A:Reference number: A90180; MUID:72243008; PMID:4625868

A:Accession: A90180

A:Molecule type: protein

A:Residues: 20-74, 'ND', '77-261 <AND>

R:Lin, K.T.D.; Deutsch, H.F.

J. Biol. Chem. 248, 1885-1893, 1973

A:Title: Human carbonic anhydrases. XI. The complete primary structure of carbonic anhyd

A:Reference number: A92128; MUID:73134579; PMID:4632246

A:Accession: A92128

A:Molecule type: protein

A:Residues: 12-26, 'DQN', '30-165, 'E', '167-261 <LIN>

A>Note: this sequence has been revised in reference A92147

R:Lin, K.T.D.; Deutsch, H.F.

J. Biol. Chem. 249, 2329-2337, 1974

A:Title: Human carbonic anhydrases. XII. The complete primary structure of the C isozyme

A:Reference number: A92147; MUID:74143468; PMID:4207120

A:Accession: B92147

cell surface-bind
hypothetical prote
carbonic anhydrase
C9 protein - rabbi
carbonate dehydrat
carbonate dehydrat
carbonate dehydrat
hypothetical prote
p54/58N - human
hypothetical prote
protein F54D8.4 [i
carbonate anhydrase
carbonate dehydrat
carbonic anhydrase
carbonic anhydrase

30 386 29.9 304 2 E36847
31 386 29.9 304 2 T28536
32 381 29.6 304 2 T37381
33 374.5 29.1 285 2 A36819
34 369.5 28.7 307 2 A29993
35 364.5 28.3 319 2 S71877
36 362.5 28.1 308 1 CRHU6
37 357 27.7 334 2 T16772
38 335.5 26.0 459 2 I38013
39 330.5 25.6 324 2 T34196
40 329 25.5 246 2 T16575
41 320.5 24.9 365 2 F88449
42 312.5 24.2 328 2 JE0375
43 304 23.6 312 1 CRHU4
44 302 23.4 264 2 AB2172
45 300.5 23.3 275 2 H83694

A;Molecule type: protein
A;Residues: 2-74, ND, 77-261 <I12>
R;Kannan, K.K.; Notstrand, B.; Fridborg, K.; Lovgren, S.; Ohlsson, A.; Petef, M.
Proc. Natl. Acad. Sci. U.S.A. 72, 51-55, 1975
A;Title: Crystal structure of human erythrocyte carbonic anhydrase B. Three-dimensional
A;Reference number: A33803; MUID:75120492; PMID:804171
A;Contents: annotation; X-ray crystallography, 2.2 angstroms
C;Genetics:
A;Gene: GDB:CA1
A;Cross-references: GDB:119047; OMIM:114800
A;Map position: 8q13-8q22.1
A;Introns: 13/1; 79/1; 118/3; 150/3; 171/3; 223/3
A;Note: the first intron occurs before the initiator codon
C;Function:
A;Description: catalyzes the reversible dissociation of carbonic acid to carbon dioxide
A;Note: this form is predominantly expressed in erythrocytes
C;Superfamily: carbonate dehydratase; carbonic anhydrase homology
C;Keywords: acetylated amino end; carbon-oxygen lyase; hydro-lyase; metalloprotein; zinc
F;2-261/Product: carbonate dehydratase I #status experimental <MAT>
F;6-261/Domain: carbonic anhydrase homology <CAH>
F;2/Modified site: acetylated amino end (Ala) (in mature form) #status experimental
F;95,97,120/Binding site: zinc (His) #status experimental
Query Match 62.4%; Score 804; DB 1; Length 261;
Best Local Similarity 59.4%; Pred. No. 1.3e-60;
Matches 142; Conservative 44; Mismatches 53; Indels 0; Gaps 0;
QY 1 MSRLSWGVEHNGPIHWKEFFPIADGQOSPIETKTEKVDSSRLPLSIKYDPSSAKTI 60
Db 1 MASPDWGYDDKNGPEQWSKLYPIANGNQSPVDIKTSEKHTSLKPISVYNPATAKEI 60
QY 61 SNSGHSFNVDFTDENKSVLRGGPLTGSYRLRQVHLHWSADDSHGEHIVDGSVYAEELH 120
Db 61 INVGHSEHVNFEEDNDRSVLRGGPFSYRLFOHFHFWGSEHTEHSGSEHTVDGVKYSAEELH 120
QY 121 VVHNSDKYPSFVEAAHEPDGLAVLGVLQIGEPNSLOKLTDTLDSIKEKGKQTRFTNF 180
Db 121 VVHNSDKYPSFVEAAHEPDGLAVLGVLQIGEPNSLOKLTDTLDSIKEKGKQTRFTNF 180
QY 181 DPSTLLPSSLDFTWYTPGSLTHPPLESYVTWICKESISVSSEQLAQFRSLLSNVEGDNA 239
Db 181 DPSTLLPSSLDFTWYTPGSLTHPPLESYVTWICKESISVSSEQLAQFRSLLSNVEGDNA 239
Query Match 62.4%; Score 804; DB 2; Length 261;
Best Local Similarity 59.8%; Pred. No. 1.3e-60;
Matches 143; Conservative 43; Mismatches 53; Indels 0; Gaps 0;
QY 1 MSRLSWGVEHNGPIHWKEFFPIADGQOSPIETKTEKVDSSRLPLSIKYDPSSAKII 60
Db 1 MASPDWGYDDKNGPEQWSKLYPIANGNQSPVDIKTSEKHTSLKPISVYNPATAKEI 60

QY 61 SNSGHSFNVDFTDENKSVLRGGPLTGSYRLRQVHLHWSADDSHGEHIVDGSVYAEELH 120
Db 61 INVGHSEHVNFEEDNDRSVLRGGPFSYRLFOHFHFWGSEHTEHSGSEHTVDGVKYSAEELH 120
QY 121 VVHNSDKYPSFVEAAHEPDGLAVLGVLQIGEPNSLOKLTDTLDSIKEKGKQTRFTNF 180
Db 121 LTHWSAKYSSLAEEAASKADGLAVIGVLMKVGSEANPKLQKILDAQLAKTKGKRAPFTNF 180
QY 181 DLLSLLPPSWDYTYTPGSLTPVPLLESYVTWIKVLPINISSQOLAKFRSLLCATGEAA 239
Db 181 DPSTLLPSSLDFTWYTPGSLTHPPLESYVTWICKESISVSSEQLAQFRSLLSNVEGDNA 239
RESULT 3
CHROID
carbonate dehydratase (EC 4.2.1.1) I - horse
N;Alternate names: carbonic anhydrase I
C;Species: Equus caballus (domestic horse)
C;Date: 22-May-1981 #sequence_revision 22-May-1981 #text_change 11-Nov-1996
C;Accession: A01140
R;Jabusch, J.R.; Bray, R.P.; Deutsch, H.F.
J. Biol. Chem. 255, 9196-9204, 1980
A;Title: Sequence of the low activity equine erythrocyte carbonic anhydrase and delineate
A;Reference number: A01140; MUID:81006999; PMID:6773961
A;Contents: D isozyme
A;Accession: A01140
A;Molecule type: Protein
A;Residues: 1-260 <JAB>
A;Note: 65-Gly, 115-His, 157-Gly, 212-Tyr, and 224-Ala were also found in the electrophore
1 isozyme has 183-Arg; and the B isozyme has 183-Arg and 222-Arg
C;Superfamily: carbonate dehydratase; carbonic anhydrase homology
C;Keywords: carbon-oxygen lyase; hydro-lyase; zinc
F;5-260/Domain: carbonic anhydrase homology <CAH>
F;94,96,119/Binding site: zinc (His) #status predicted
Query Match 62.1%; Score 800; DB 1; Length 260;
Best Local Similarity 62.0%; Pred. No. 2.8e-60;
Matches 145; Conservative 35; Mismatches 54; Indels 0; Gaps 0;
QY 6 WGYREHNGPIHWKEFFPIADGQOSPIETKTEKVDSSRLPLSIKYDPSSAKIISNGH 65
Db 5 WGYDPSBGPZEWVKLYPIABGBQSPIDIKTSEKHTSLKPSVSYDPATAKEIVNVGH 64
QY 66 SFNVDPDDTENKSVLRGGPLTGSYRLRQVHLHWSADDSHGEHIVDGSVYAEELHVVHVN 125
Db 65 SPQVKFEDSDNRSVLRKDGPLPGSYRLVQPFHFWGSEHTDYGSEHTVDGVKYSAEELHVVHVN 124
QY 126 SDKYPSEVFAAHEPDGLAVLGVLQIGEPNSLOKLTDTLDSIKEKGKQTRFTNF 185
Db 125 SSKYSSFDEASSQADGLAILGLVMKVGSEANPKLQKVLDAALNEVTKTGKAPFKNDFPSSL 184
QY 186 LPSSDYTYTPGSLTPVPLLESYVTWIKVLPINISSQOLAKFRSLLCATGEAA 239
Db 185 LPSSDYTYTPGSLTHPPLESYVTWICKENISISSQOLSQFRSLLSNVEGGKA 238
RESULT 4
JN0835
carbonate dehydratase (EC 4.2.1.1) I - chimpanzee
N;Alternate names: carbonic anhydrase
C;Species: Pan troglodytes (chimpanzee)
C;Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 07-May-1999
C;Accession: JN0835
R;Epperly, B.R.; Bergenhem, N.C.H.; Venta, P.J.; Tashian, R.E.
Gene 131, 249-253, 1993
A;Title: Characterization of the genes encoding carbonic anhydrase I of chimpanzee and
A;Reference number: JN0835; MUID:94010316; PMID:8406018
A;Accession: JN0835
A;Molecule type: mRNA
A;Residues: 1-261 <EPP>
A;Cross-references: GB:L11621
A;Note: the authors translated codon GAG for residue 206 as Leu

DB 183 LKRALNYWTYPGSLTNPPLESVTWWLKEPTSVSSQQLKFRSLNFNAEGE 234

A;Contents: annotation; X-ray crystallography, 2.0 angstroms, residues 4-259
R.Liljas, A.; Kanner, K.K.; Bergsten, P.C.; Waara, I.; Fridborg, K.; Strandberg, B.; Carl
Nature New Biol. 235, 131-137, 1972
A;Title: Crystal structure of human carbonic anhydrase C.
A;Reference number: A93404; MUID:72111787; PMID:4621826
A;Contents: annotation; X-ray crystallography, 2.0 angstroms
A;Note: Other residues at the active site are His-64, Asn-67, Tyr-127, Leu-197, Thr-198,
C;Genetics:
A;Gene: GDB:CA2
A;Cross-references: GDB:119739; OMIM:259730
A;Map position: 8q13-8q22-1
A;Introns: 12/1; 78/1; 117/3; 148/3; 169/3; 221/3
C;Function:
A;Description: catalyzes the reversible dissociation of carbonic acid to carbon dioxide
A;Note: this form is expressed in erythrocytes and other tissues; deficiency of this for
C;Superfamily: carbonate dehydratase; carbonic anhydrase homology
C;Keywords: acetylated amino end; carbon-oxygen lyase; hydro-lyase; metalloprotein; mono
F;2-260/Product: carbonate dehydratase II #status experimental <MAT>
F;5-259/Domain: carbonic anhydrase homology <CAH>
F;5-259/Modified site: acetylated amino end (Ser) (in mature form) #status experimental
F;2/Domain site: acetylated amino end (Ser) #status experimental
F;94,96,119/Binding site: zinc (His) #status experimental

Query Match 61.2%; Score 788.5; DB 1; Length 260;
Best Local Similarity 61.6%; Pred. No. 2.6e-59;
Matches 143; Conservative

Qy 6 WGYREHNGPTHWKEFFPIADGQOSPIETKTEVKYDYSLSRPLSIKYDPSSAKLIISNGH 65
Db 5 WGYGKHNGPEHWHKDFPIAKGQSPVIDIDHTAKYDPSLKPLSVSDQATSLRIILNGH 64
Qy 66 SENVDFFDTENKSVLRGGPLTGSVRLRQVHLHWSADHGHSEHIVDGVSYAAELHVWHVN 125
Db 65 AENVEFDSDQKAVLKGGPLDGYTRLLOFHFHWSLDCQGSSEHTVDVKKYAAELHLVWHVN 124
Qy 126 SDKYFSEVAAHEPDGLAVLGVLQIGEPNSQLQKITDILDSIKBKQKQTRFNFEDLLSL 185
Db 125 T-KYGDGKAVQQDGLAVLGIFLKGSAKPGQLQKVVDLVDSIKTKGKSADFTNFDPRGL 183
Qy 186 LPPSDWTYTPGSLTVPPLLESVTWIVLKOPINISSQOLAKFRSLICTAEGE 237
Db 184 LPESLDWTYTPGSLTTPLEECVTWIVLKGPISVSSQVLKFRKLNPNNGEGE 235

RESULT 7

CRS02
carbonate dehydratase (EC 4.2.1.1) II - bovine (tentative sequence)
N;Alternate names: carbonic anhydrase II
C;Species: Bos primigenius taurus (cattle)
C;Date: 31-Mar-1981 #sequence_revision 31-Mar-1981 #text_change 31-Mar-2000
C;Accession: A01144
R;Schlaky, M.; Limozin, N.; Filippi-Foveau, D.; Gullian, J.M.; Laurent-Tabusse, G.
Biochimie 58, 1071-1082, 1976
A;Title: Structure primaire de l'anhydrase carbonique erythrocytaire bovine Cl. II. - Se
A;Reference number: A90669; MUID:77065798; PMID:826282
A;Accession: A01144
A;Molecule type: protein
A;Residues: 1-259 <SCI>
R;Gullian, J.M.; Limozin, N.; Mallet, B.; Di Costanzo, J.; Charrel, M.
Biochimie 59, 293-302, 1977
A;Title: Independance genetique de deux formes de l'anhydrase carbonique erythrocytaire
A;Reference number: A90672; MUID:77242599; PMID:19093
A;Contents: annotation
A;Note: one minor and two major forms were isolated chromatographically. One of the major
C;Superfamily: carbonate dehydratase; carbonic anhydrase homology
C;Keywords: acetylated amino end; carbon-oxygen lyase; hydro-lyase; zinc
F;4-258/Domain: carbonic anhydrase homology <CAH>
F;1/Modified site: acetylated amino end (Ser) #status experimental
F;93,95,118/Binding site: zinc (His) #status predicted

Query Match 60.9%; Score 785.5; DB 1; Length 259;
Best Local Similarity 61.2%; Pred. No. 4.7e-59;
Matches 142; Conservative 41; Mismatches 48; Indels 1; Gaps 1;

F:5-260/Domain: carbonic anhydrase homology <CAH>
F:1/Modified site: acetylated amino end (Ala) #status experimental
F:94,96,119/Binding site: zinc (His) #status predicted

Query Match 60.5%; Score 780; DB 1; Length 260;
Best Local Similarity 58.9%; Pred. No. 1.4e-58;
Matches 136; Conservative 46; Mismatches 49; Indels 0; Gaps 0;

QY 6 WGYREHNGPIHWKEFFPIADGQOQSPIEIKTEVKYDSSLRPLSIKYDPSSAKIISNSGH 65
DB 5 WGYDDKNGPEQWSKLYPIANGNQSPDIKTSEAKHDTSLKPTISVYNPATAKEIINVGH 64
QY 66 SFNVDFDDTENKSVLRGGPLTGSYRLQVHLHWGSADHDGSEHIVDGVSYAAELHVVHWN 125
DB 65 SFHVFEDNDKRSVLKGGPFSOSYRUFQFHHWGSNEYSGSEHTVDGVKYSSELHVVHWN 124
QY 126 SDKYPFVEAAHEPDPGLAVLGVLQIGEPNSOLQKITDITLDSIKEKGKQTRFTNFDLLSL 185
DB 125 SAKYSLSAEAVSKADGLAVIGVLMKVGEANPKLQKVLDAHLKTKGKRAPFTNFDPSLT 184
QY 186 LPSPWDYWTYPGSLTPVPPLESVTWVLKQPINISSQQLAKFRSLICTAEG 236
DB 185 LPSSLDFTWYSGSLTHPPLYESVTWLTICKESISVSSEQLAQRSLLSNVEG 235

RESULT 10
JC2580
carbonate dehydratase (EC 4.2.1.1) II - chicken
N:Alternate names: carbonic anhydrase II
C:Species: Gallus gallus (chicken)
C:Date: 13-Jun-1995 #sequence revision 14-Jul-1995 #text change 20-Jun-2000
A:Accession: JC2580; S10229; S01078; A26415; I50181; S31987
R:Merzquita, W.; Pau, M.; Mezquita, C.
Gene 147, 231-235, 1994
A:Title: A novel carbonic anhydrase II mRNA isolated from mature chicken testis displays
A:Reference number: JC2580; MUID:95011620; PMID:7926806
A:Accession: JC2580
A:Molecule type: mRNA
A:Residues: 1-260 <MEZ>
A:Cross-references: EMBL:X17378; NID:G65331; PIDN:CAA78681.1; PID:G65332
A:Experimental source: testis
R:Godbout, R.; Anderson, R.; Upton, C.; Day, R.
Nucleic Acids Res. 18, 1049, 1990
A:Title: Utilization of the second polyadenylation signal at the 3' end of the chicken c
A:Reference number: S10229; MUID:90192090; PMID:1969140
A:Accession: S10229
A>Status: translation not shown
A:Molecule type: mRNA
A:Residues: 222-260 <GOD>
A:Cross-references: EMBL:X17378; NID:G63127; PIDN:CAA35250.1; PID:G63128
R:Yoshihara, C.M.; Lee, J.D.; Dodgson, J.B.
Nucleic Acids Res. 15, 753-770, 1987
A:Title: The chicken carbonic anhydrase II gene: evidence for a recent shift in intron p
A:Reference number: S01078; MUID:87146391; PMID:3029691
A:Accession: S01078
A:Molecule type: DNA
A:Residues: 1-249, 'V', 251-260 <YOS>
A:Cross-references: EMBL:X06000; NID:G63115; PIDN:CAA29417.1; PID:G1289219
A:Note: the authors translated the codon GTA for residue 250 as Leu
R:Rogers, J.H.
Eur. J. Biochem. 162, 119-122, 1987
A:Title: Sequence of carbonic anhydrase II cDNA from chick retina.
A:Reference number: A26415; MUID:87133522; PMID:3102231
A:Accession: A26415
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 'L', 6-260 <ROG>
A:Cross-references: EMBL:X04810; NID:G63129; PIDN:CAA28501.1; PID:G833606
R:Yoshihara, C.M.; Federspiel, M.; Dodgson, J.B.
Ann. N. Y. Acad. Sci. 429, 332-334, 1984
A:Title: Isolation of the chicken carbonic anhydrase II gene.
A:Reference number: I50181; MUID:84255154; PMID:6331256
A:Accession: I50181

A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 'GX', 10-16, 'X', 18, 'X', 20-43, 'X', 45-78, 'X', 80-86, 'S' <YO2>
A:Cross-references: GB:M25943; NID:G211369; PIDN:AAA48646.1; PID:G211370
C:Genetics:
A:Introns: 12/1; 78/1; 117/3; 148/3; 169/3; 221/3
C:Function:
A:Description: catalyzes the reversible dissociation of carbonic acid to carbon dioxide
C:Superfamily: carbonate dehydratase; carbonic anhydrase homology
C:Keywords: carbon-oxygen lyase; hydro-lyase; metalloprotein; zinc
F:5-259/Domain: carbonic anhydrase homology <CAH>
F:94,96,119/Binding site: zinc (His) #status predicted

Query Match 60.4%; Score 778.5; DB 2; Length 260;
Best Local Similarity 60.3%; Pred. No. 1.8e-58;
Matches 140; Conservative 30; Mismatches 61; Indels 1; Gaps 1;

QY 6 WGYREHNGPIHWKEFFPIADGQOQSPIEIKTEVKYDSSLRPLSIKYDPSSAKIISNSGH 65
DB 5 WGYDSHNGPAHWEHFPPIANGERSQSPIALSTKAARDIPALKEPLSFSDGTAKAIVNNGH 64
QY 66 SFNVDFDDTENKSVLRGGPLTGSYRLQVHLHWGSADHDGSEHIVDGVSYAAELHVVHWN 125
DB 65 SFNVEFDDSSDKSVLQGGALDGVYRLVQPHIHWGSCGQSGSEHTVDGVKYDAELHVVHN 124
QY 126 SDKYPFVEAAHEPDPGLAVLGVLQIGEPNSOLQKITDITLDSIKEKGKQTRFTNFDLLSL 185
DB 125 V-KYGFKAELKHPDGLAVVGIFMKVGNAPKIQKVDALNSIQTKGQASFTNFDPTGL 183
QY 186 LPSPWDYWTYPGSLTPVPPLESVTWVLKQPINISSQQLAKFRSLICTAEG 237
DB 184 LPPCRDIWTYPGSLTTPPLHECVIHWLKEPITVSSEOMCKLUGLCFSAENE 235

RESULT 11
CRRB?

carbonate dehydratase (EC 4.2.1.1) II - rabbit (tentative sequence)
N:Alternate names: carbonic anhydrase II
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 31-Mar-1981 #sequence revision 31-Mar-1981 #text change 31-Mar-2000
A:Accession: A01142
R:Perrelli, R.E.; Stroup, S.K.; Tanis, R.J.; Tashian, R.E.
Biochim. Biophys. Acta 533, 1-11, 1978
A:Title: Amino acid sequence of rabbit carbonic anhydrase II.
A:Reference number: A01142; MUID:78144871; PMID:416851
A:Accession: A01142
A:Molecule type: protein
A:Residues: 1-259 <PER>
A:Note: 203-Glu was also found
C:Superfamily: carbonate dehydratase; carbonic anhydrase homology
C:Keywords: acetylated amino end; carbon-oxygen lyase; hydro-lyase; zinc
F:4-258/Domain: carbonic anhydrase homology <CAH>
F:1/Modified site: acetylated amino end (Ser) #status experimental
F:93,95,118/Binding site: zinc (His) #status predicted

Query Match 59.1%; Score 761.5; DB 1; Length 259;
Best Local Similarity 60.8%; Pred. No. 5.1e-57;
Matches 141; Conservative 31; Mismatches 59; Indels 1; Gaps 1;

QY 6 WGYREHNGPIHWKEFFPIADGQOQSPIEIKTEVKYDSSLRPLSIKYDPSSAKIISNSGH 65
DB 4 WGYGHNGPEHWHKDFPIADGEROSPIDIDTDAAKDPSLKLPLRVSYEHPISRIINNGH 63
QY 66 SFNVDFDDTENKSVLRGGPLTGSYRLQVHLHWGSADHDGSEHIVDGVSYAAELHVVHWN 125
DB 64 SFNVEFDDSHDKSVLKEGPLEGTLYLIQFHHWGSDDGEGSEHTVKKKYAAELHVVHN 123
QY 126 SDKYPFVEAAHEPDPGLAVLGVLQIGEPNSOLQKITDITLDSIKEKGKQTRFTNFDLLSL 185
DB 124 T-KYDGFKA VKHPDGLAVLGIFLKG SATPGQKVVDTLSSIKTKGKSVDFTNFDPRGL 182
QY 186 LPSPWDYWTYPGSLTPVPPLESVTWVLKQPINISSQQLAKFRSLICTAEG 237

Db 183 LPESLDYWTYPGSLTPPQLQCVTWVILKEPITVSSQMLKFRNLNFNKEAE 234

RESULT 12

A22612

carbonate dehydratase (EC 4.2.1.1) III - horse

N;Alternate names: carbonate dehydratase III

C;Species: Equus caballus (domestic horse)

C;Date: 22-Jul-1987 #sequence_revision 22-Jul-1987 #text_change 25-Apr-1997

C;Accession: A22612

R;Wendorff, K.M.; Nishita, T.; Jabusch, J.R.; Deutsch, H.F.

J. Biol. Chem. 260, 6129-6132, 1985

A;Title: The sequence of equine muscle carbonic anhydrase.

A;Reference number: A22612; MUID:85207593; PMID:3922970

A;Accession: A22612

A;Molecule type: protein

A;Residues: 1-259 <WEN>

C;Superfamily: carbonate dehydratase; carbonic anhydrase homology

C;Keywords: acetylated amino end; carbon-oxygen lyase; hydro-lyase; zinc

F;4-258/Domain: carbonic anhydrase homology <CAH>

F;1/Modified site: acetylated amino end (Ala) #status predicted

Query Match 58.5%; Score 754.5; DB 2; Length 259;

Best Local Similarity 59.1%; Pred. No. 2e-56;

Matches 137; Conservative 30; Mismatches 64; Indels 1; Gaps 1;

QY 6 WGYREHNGPIHWKEFPPIADGQQSPDIEIKTEVKYDSSLRPLSLKYDPSSAKIISNSGH 65

DB 4 WGYADHNGPDHWEFFYPIAKGNQSPIELHTKDINHPSLKAWTASYDPGSAKTLNNGR 63

QY 66 SRNVDPDDTENKSVLRGGPLTGSYRLRQVHLHWGSADDSHSEHIVDGVSYAAELHVVHVN 125

DB 64 TCRVVEDDYYDRSMLRGGLTAPYRLRQPLHLHWGSSDDHSGSEHTVDGVKYAAELHVLHVN 123

QY 126 SDKYPSEFVAEAHPDGLAVLGVFLQIGEPNSLOKLTITDLSIKKKGKOTRFTNFDLLSL 185

DB 124 -PRYNTYGGALQPDGIAVGVFLKIGREKGEPLFDALDKIKYKGAAPFNFDPSCL 182

QY 186 LPSPSWDYTPGSLTPVPLLESVTWIVLKOPINISSQQLAKFRSLCTAAGE 237

DB 183 FPTCRDYWTYRGSTFTTPCEBICVLLKKEPITVSDSQVAKRLSFSAAENE 234

RESULT 13

A26344

carbonate dehydratase (EC 4.2.1.1) I - mouse

C;Species: Mus musculus (house mouse)

C;Date: 21-May-1988 #sequence_revision 21-May-1988 #text_change 22-Jun-1999

C;Accession: A26344; I49573

R;Fraser, P.J.; Curtis, P.J.

J. Mol. Evol. 23, 294-299, 1986

A;Title: Molecular evolution of the carbonic anhydrase genes: calculation of divergence

A;Reference number: A26344; MUID:87169766; PMID:3104601

A;Accession: A26344

A;Molecule type: mRNA

A;Residues: 1-261 <FRA>

R;Fraser, P.; Cummings, P.; Curtis, P.

Mol. Cell. Biol. 9, 3308-3313, 1989

A;Title: The mouse carbonic anhydrase I gene contains two tissue-specific promoters.

A;Reference number: I49573; MUID:90014784; PMID:2571923

A;Accession: I49573

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-82, 'G', 83-100, 102-261 <RES>

A;Cross-references: GB:L36655; NID:g556293; PIDN:AAA50291.1; PID:g556295

C;Genetics:

A;Gene: Car1

A;Introns: 13/1; 79/1; 118/3; 150/3; 171/3; 223/3

C;Superfamily: carbonate dehydratase; carbonic anhydrase homology

C;Keywords: carbon-oxygen lyase; hydro-lyase; zinc

F;6-261/Domain: carbonic anhydrase homology <CAH>

Query Match 58.5%; Score 754; DB 2; Length 261;

Best Local Similarity 59.2%; Pred. No. 2.2e-56;

Matches 142; Conservative 36; Mismatches 60; Indels 2; Gaps 2;

QY 1 MSRLSWGYREHNGPIHWKEFPPIADGQQSPDIEIKTEVKYDSSLRPLSLKYDPSSAKII 60

DB 1 MASADWGYSGENGPQWSKLYPIANGNNQSPDIKTSEAHNDSSLKPLSLISYNPATAKEI 60

QY 61 SNSGHSFNVDPTENKSVLRGGPLTGSYRLRQVHLHWG-SADDSHSEHIVDGVSYAAEL 119

DB 61 VNVGHSFHVIFDDSSNQSVLK-GPLADSYRLTFQHFHMGNSNDHSGSEHTVDGTRYSGEL 119

QY 120 HVVHWNSDKYPSFVEAHPDGLAVLGVFLQIGEPNSLOKLTITDLSIKKKGKOTRFTN 179

DB 120 HLHVHNSAKYSASASAIKADGLAILGLVMKVGPNPSLQKVLDAI NSVTKTKRAPFTN 179

QY 180 FDLSSLPPSWDYTPGSLTPVPLLESVTWIVLKOPINISSQQLAKFRSLCTAAGEAA 239

DB 180 FDPSSLPSLDYWTYFGSLTHPPLHESVTWVICKDSLSLPEQLAQLRGLLSAAGESA 239

RESULT 14

CRMS2

carbonate dehydratase (EC 4.2.1.1) II - mouse

N;Alternate names: carbonic anhydrase II

C;Species: Mus musculus (house mouse)

C;Date: 25-Feb-1985 #sequence_revision 06-Feb-1995 #text_change 18-Jun-1999

C;Accession: A23900; B23202; A01143; A20539; I51949

R;Venta, P.J.; Montgomery, J.C.; Hewett-Emmet, D.; Wiebauer, K.; Tashian, R.E.

J. Biol. Chem. 260, 12130-12135, 1985

A;Title: Structure and exon to protein domain relationships of the mouse carbonic anhydrase

A;Reference number: A23900; MUID:86008276; PMID:2995362

A;Accession: A23900

A;Molecule type: DNA

A;Residues: 1-260 <VEN>

A;Experimental source: strain YER

R;Venta, P.J.; Montgomery, C.; Hewett-Emmett, D.; Tashian, R.E.

Biochim. Biophys. Acta 826, 195-201, 1985

A;Title: Comparison of the 5' regions of human and mouse carbonic anhydrase II genes and

A;Reference number: A90655; MUID:86077780; PMID:3000449

A;Accession: B23202

A;Molecule type: DNA

A;Residues: 1-77 <VE2>

A;Note: the authors translated the codon CAG for residue 39 as His

R;Curtis, P.J.; Withers, E.; Demuth, D.; Watt, R.; Venta, P.J.; Tashian, R.E.

Gene 25, 325-332, 1983

A;Title: The nucleotide sequence and derived amino acid sequence of cDNA coding for mouse

A;Reference number: A01143; MUID:84109569; PMID:6420240

A;Accession: A01143

A;Molecule type: mRNA

A;Residues: 2-38, 'H', 40-260 <CUR>

A;Cross-references: GB:K00811; GB:K00812; GB:M11830; NID:g192333; PIDN:AAA37356.1; PID:g

A;Note: Initiator Met not shown

R;Curtis, P.J.

J. Biol. Chem. 258, 4459-4463, 1983

A;Title: Cloning of mouse carbonic anhydrase mRNA and its induction in mouse erythroleuk

A;Reference number: A20539; MUID:83161023; PMID:6187736

A;Accession: A20539

A;Molecule type: mRNA

A;Residues: 155-178; 214-240 <CU2>

R;Venta, P.J.; Montgomery, J.C.; Wiebauer, K.; Hewett-Emmett, D.; Tashian, R.E.

Ann. N. Y. Acad. Sci. 429, 309-323, 1984

A;Title: Organization of the mouse and human carbonic anhydrase II genes.

A;Reference number: I51949; MUID:84255152; PMID:6331255

A;Accession: I51949

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 241-260 <RES>

A;Cross-references: GB:M25944; NID:g199078; PIDN:AAA39505.1; PID:g199079

C;Genetics:

A;Gene: Car-2

A;Introns: 12/1; 78/1; 117/3; 144/1; 169/3; 221/3

C;Superfamily: carbonate dehydratase; carbonic anhydrase homology

C;Keywords: carbon-oxygen lyase; hydro-lyase; zinc

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 14:58:23 ; Search time 124 Seconds
(without alignments)
551.423 Million cell updates/sec

Title: US-10-069-434-1

Perfect score: 1289

Sequence: 1 MSRLSWGVRHNGPIHWKEF.....QLAKFRSLCTAEGEAAFL 242

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq 29Jan04:*

1: Geneseqp1980s:*

2: Geneseqp1990s:*

3: Geneseqp2000s:*

4: Geneseqp2001s:*

5: Geneseqp2002s:*

6: Geneseqp2003as:*

7: Geneseqp2003bs:*

8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1289	100.0	242	5	ABb08900 Human lya
2	1289	100.0	262	6	ABj37886 NOVX proc
3	1289	100.0	274	4	AAu19418 Human dia
4	1288	99.9	262	6	ABj37885 NOVX prot
5	1193.5	92.6	247	5	AAe25377 Human NZM
6	864	67.0	184	3	AAb63110 Human sec
7	804	62.4	261	5	AAo15236 Human car
8	804	62.4	261	6	ADa10962 Human cDN
9	802	62.4	263	4	AAg73863 Human col
10	802	62.4	260	4	AAb59588 Human car
11	791.5	61.4	260	2	AAw75702 Carbonic
12	788.5	61.2	259	4	AAb59589 Human car
13	788.5	61.2	259	7	ADe62800 Human Pro
14	788.5	61.2	259	7	ADe63735 Human Pro
15	788.5	61.2	260	6	ADa10999 Human cDN
16	788.5	61.2	288	2	AAw23378 Fusion co
17	788.5	61.2	294	3	AAb53405 Human col
18	761.5	59.1	259	7	ADe54976 Rat Prote
19	750	58.2	143	4	AAo12529 Human pol
20	748.5	58.1	259	4	AAb59590 Human car
21	748.5	58.1	259	7	ADe54978 Human Pro
22	748.5	58.1	259	7	ADe62798 Rat Prote
23	748.5	58.1	259	7	ADe63733 Rat Prote
24	748.5	58.1	260	5	ABb57256 Mouse isc
25	747.5	58.0	260	5	ABp62813 Human pol

26	747.5	58.0	421	5	AAm51157	Streptavi
27	719.5	55.8	242	5	ABp69520	Human pol
28	705.5	54.7	256	6	ABU11700	Human MDD
29	695	53.9	201	2	AAW06552	Human col
30	695	53.9	201	2	AAW46884	Protein s
31	695	53.9	201	5	AAU76212	Human col
32	695	53.9	201	5	ABG32041	Human col
33	668	51.8	261	4	AAb59593	Human car
34	660	51.2	329	4	AAU30162	Novel hum
35	589	45.7	323	4	ABG12955	Novel hum
36	577.5	44.8	183	3	AAb63109	Gene 27 h
37	547	42.4	270	4	ABb63491	Drosophil
38	451	35.0	289	5	ABJ05479	Human bre
39	413.5	32.1	274	2	AAr91953	Lung can
40	411.5	31.9	270	2	AAr91952	Lung can
41	411.5	31.9	325	2	AAr91951	Lung can
42	411.5	31.9	325	2	AAr91955	HCAVIIII p
43	411.5	31.9	343	6	ABr58589	Human can
44	411.5	31.9	354	2	AAr91950	Lung can
45	411.5	31.9	354	3	AAy96200	Non-small

ALIGNMENTS

RESULT 1
ABb08900

ID ABb08900 standard; protein; 242 AA.

XX AC ABb08900;

XX DT 18-JUN-2002 (first entry)

XX DE Human lyase HLVA-1 protein.

XX KW Cytostatic; anticonvulsant; cerebroprotective; neurotropic; virucide;

XX KW neuroprotective; antibacterial; antidiabetic; antiinflammatory; antigout;

XX KW ophthalmological; hypotensive; immunosuppressive; dermatological;

XX KW nephrotrophic; antithyroid; thyromimetic; osteopathic; antipsoriatic;

XX KW antitumor; fungicide; antiparasitic; protozoacide; tranquilizer; cancer;

XX KW neuroleptic; diagnosis; treatment; immunological disorder; AIDS; allergy;

XX KW acquired immunodeficiency syndrome; asthma; HLVA; infection; anaemia;

XX KW Crohn's disease; multiple sclerosis; atherosclerosis; osteoporosis;

XX KW rheumatoid arthritis; Alzheimer's; Parkinson's disease; epilepsy; stroke;

XX KW muscular dystrophy; Down's syndrome; myaesthesia gravis; glaucoma;

XX KW transgenic; gene therapy; drug screening; human lyase; enzyme.

XX OS Homo sapiens.

XX PN WO200200840-A2.

XX PD 03-JAN-2002.

XX PF 13-JUN-2001; 2001WO-US019166.

XX PR 23-JUN-2000; 2000US-0213383P.

XX PR 30-JUN-2000; 2000US-0215544P.

XX PR 04-AUG-2000; 2000US-0222818P.

XX PA (INCY-) INCYTE GENOMICS INC.

XX PI Thornton M, Ramkumar J, Tribouley CM, Yue H, Nguyen DB, Yao MG;

XX PI Patterson C, Gandhi AR, Burford N, Thangavelu K, Baughn MR;

XX DR WPI; 2002-139910/18.

XX DR N-PSDB; ABA97691.

XX PT New isolated human lyase polypeptide for diagnosing, treating and

XX PT preventing e.g. glaucoma, ocular hypertension, stroke, asthma, or gout.

XX PS Claim 1; Page 96-97; 101pp; English.

XX CC The present sequence represents human lyase polypeptide (HLVA-1) encoded

CC presence of DITHPs in samples (e.g. by enzyme linked immunosorbant assay
CC (ELISA)). AAU19415-AAU19625 represent human diagnostic and therapeutic
CC (DITHP) polypeptides of the invention
XX
SQ Sequence 274 AA;

Query Match 100.0%; Score 1289; DB 4; Length 274;
Best Local Similarity 100.0%; Pred. No. 7.3e-126;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSRLSGVYEHNGPIHWKEFFPIADGQDSPIETIKTEVKYDSSLRPLSIKYDPSAKII 60
Db |||||
13 MSRLSGVYEHNGPIHWKEFFPIADGQDSPIETIKTEVKYDSSLRPLSIKYDPSAKII 72
QY 61 SNSGHSFNVDFTDNTENKSVLRGPGLTGSYRLRQVHLHWSADHSGSEHIVDGVSYAAELH 120
Db |||||
73 SNSGHSFNVDFTDNTENKSVLRGPGLTGSYRLRQVHLHWSADHSGSEHIVDGVSYAAELH 132
QY 121 VVHNSDKYPSVEAAHEPDGLAVLGFLQIGEPNSQLQKITDTLDSIKEKGKQTRFTNF 180
Db |||||
133 VVHNSDKYPSVEAAHEPDGLAVLGFLQIGEPNSQLQKITDTLDSIKEKGKQTRFTNF 192
QY 181 DLLSLPPSWDYWTYPGSLTVPPLESVTWVLKQPINISSQOLAKFRSLLCTAEGEAAA 240
Db |||||
193 DLLSLPPSWDYWTYPGSLTVPPLESVTWVLKQPINISSQOLAKFRSLLCTAEGEAAA 252
QY 241 FL 242
Db ||
253 FL 254

RESULT 4
ABJ37885
ID ABJ37885 standard; protein; 262 AA.
XX
AC ABJ37885;
XX
DT 22-MAY-2003 (first entry)
XX
DE NOVX protein sequence SEQ ID No 16.
XX
KW Hepatotropic; immunosuppressive; cardiact; hypertensive; tranquilizer;
KW virucide; antibacterial; protozoacide; fungicide; nootropic;
KW antiparasitic; neuroprotective; cerebroprotective; antiparkinsonian;
KW anticonvulsant; antiaddictive; analgesic; dermatological; keratolytic;
KW antieborrheic; antirheumatic; antiarthritic; antiinflammatory; anti-HIV;
KW cytostatic; antiasthmatic; antipsoriatic; hypotensive; osteopathic;
KW antitumor; anorectic; antidiabetic; antiallergic; haemostatic;
KW neuroleptic; antidepressant; antiinfertility; NOVX; human disease;
KW NOVX-associated disorder; trauma; viral; bacterial; fungal; protozoal;
KW parasitic infection; Alzheimer's disease; stroke; forensic biology;
KW immunogen; non-human transgenic animal; gene therapy.
XX
OS Unidentified.
XX
XX
PN W0200281517-A2.
XX
XX 17-OCT-2002.
XX
XX 22-JAN-2002; 2002WO-US002064.
XX
PR 19-JAN-2001; 2001US-0262892P.
PR 23-JAN-2001; 2001US-0263598P.
PR 24-JAN-2001; 2001US-0263799P.
PR 25-JAN-2001; 2001US-0264117P.
PR 25-JAN-2001; 2001US-0264139P.
PR 26-JAN-2001; 2001US-0264478P.
PR 30-JAN-2001; 2001US-0265351P.
PR 02-MAR-2001; 2001US-0272870P.
PR 14-MAR-2001; 2001US-0275927P.
PR 14-MAR-2001; 2001US-0275930P.
PR 15-MAR-2001; 2001US-0276449P.
PR 20-MAR-2001; 2001US-0277358P.

23-MAR-2001; 2001US-0278151P.
29-MAR-2001; 2001US-0279857P.
20-APR-2001; 2001US-0285140P.
20-APR-2001; 2001US-0285141P.
30-APR-2001; 2001US-0287484P.
17-MAY-2001; 2001US-0291701P.
08-JUN-2001; 2001US-0296960P.
10-JUL-2001; 2001US-0304353P.
10-JUL-2001; 2001US-0304353P.
09-AUG-2001; 2001US-0311289P.
13-AUG-2001; 2001US-0311975P.
16-AUG-2001; 2001US-0312937P.
18-OCT-2001; 2001US-0330227P.
29-NOV-2001; 2001US-0334198P.

(CURA-) CURAGEN CORP.

Decristofaro MF, Padigaru M, Miller C, Tchernev V, Zhong H;
Zhong M, Anderson D, Ballinger R, Gerlach V, Spytek KA, Rastelli L;
Kekuda R, Guo X, Zerhusen B, Andrew D, Mezes P, Patturajan M;
Burgess CE, Eissen A, Wolenc A, Baumgartner J, Shinkets RA, Gusev V;
Vernet CAM, Taupier RJ, Pena C, Shenoy S, Li L, Casman S, Boldog F;
Fernandes E, Smithson G, Malyankar U, Tallon B, Liu X;

WPI: 2003-058504/05.
N-PSDB; ABY33350.

New polypeptides, designated as NOVX, useful for diagnosing and treating
infections, neurological diseases, cancer, allergy, and bone,
immunological, skin, renal, brain, muscle and autoimmune disorders.

Claim 1; Page 66; 672pp; English.

The invention relates to a novel isolated polypeptide, designated NOVX
(NOV1 - 33), consisting of a mature form of one of 61 sequences, given in
the specification, or its variant, where amino acid residue(s) in the
variant differ from the mature form, provided that the variant differs in
not more than 15 % of the amino acids from the sequence of the mature
form. The NOVX polypeptides, nucleic acids encoding the polypeptides, and
an antibody to the polypeptides, are useful for treating or preventing a
NOVX-associated disorder in humans and for treating a syndrome associated
with a human disease (NOVX-associated disorder). NOVX polypeptides and
the encoding nucleic acids, are useful for determining the presence of or
predisposition to a disease associated with altered levels of NOVX
polypeptide and polynucleotide, by measuring the level of polypeptide
expression or the amount of nucleic acid from a mammal and comparing it
with another mammal not having or not predisposed to the disease. NOVX
polypeptide is also useful for identifying an agent that binds to NOVX
and a cell expressing NOVX is useful for identifying an agent that
modulates the expression or activity of NOVX. The antibodies and a
polypeptide having 95 % sequence identity to NOVX polypeptide are useful
for treating a pathological state in a mammal. The antibodies are also
useful for determining the presence or amount of NOVX in a sample. NOVX
polypeptides, polynucleotides and antibodies specific for the
polypeptides are useful for treating or preventing disorders or syndromes
including trauma, viral, bacterial, fungal, protozoal, and parasitic
infections. They can also treat disorders such as e.g., Alzheimer's
disease or a stroke. The NOVX encoding nucleic acids are useful for
expressing the NOVX proteins, to detect NOVX mRNA, or a genetic lesion in
a NOVX gene and to modulate NOVX activity. NOVX sequences are also useful
for identifying a cell or tissue type in a biological sample, to amplify
DNA sequences from very small biological samples such as tissues e.g.
hair or skin or body fluids in forensic biology and as primers and probes
for use in identifying and/or cloning NOVX homologues in other cell
types. The NOVX proteins are useful as an immunogen to generate
antibodies which are useful for diagnostically monitoring protein levels
and modulating NOVX activity. Cells comprising NOVX nucleic acids are
useful for producing non-human transgenic animals which are useful for
studying the function and/or activity of NOVX protein and for identifying
and/or evaluating modulators of NOVX protein activity. The NOVX nucleic
acids can be used in gene therapy. This sequence represents a NOVX
protein of the invention

KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW cerebral ischaemia; angiogenesis; nervous system disorder; infection;
KW Alzheimer's disease; ocular disorder; corneal infection; wound healing;
KW skin aging; food additive; preservative.
XX Homo sapiens.
XX WO200061748-A1.
XX 19-OCT-2000.
XX 06-APR-2000; 2000WO-US008982.
XX 09-APR-1999; 99US-0128696P.
XX 14-JAN-2000; 2000US-0176069P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Ruben SM, Komatsoulis G;
XX WPI; 2000-638566/61.
XX New nucleic acid molecules encoding 48 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives.
XX Disclosure; Page 458-459; 480pp; English.
XX AAF22316 to AAF22363 encode the human secreted proteins given in AAB63049
CC to AAB63096. AAB63097 to AAB63132 represent more human secreted proteins
CC and polypeptides homologous to them. Human secreted proteins have
CC activities based on the tissues and cells the genes are expressed in.
CC Examples of activities include: immunosuppressive; antiarthritic;
CC antirheumatic; antiproliferative; cytostatic; cardiac; vasotropic;
CC cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide;
CC fungicide; ophthalmological; and vulnary. The polynucleotides and
CC proteins can be used to prevent, treat or ameliorate a medical condition
CC in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or
CC sheep. They are also used in diagnosing a pathological condition or
CC susceptibility to a pathological condition. Disorders which are diagnosed
CC or treated include autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi and
CC ocular disorders e.g. corneal infection. The polypeptides can also be
CC used to aid wound healing and epithelial cell proliferation, to prevent
CC skin aging due to sunburn, to maintain organs before transplantation, for
CC supporting cell culture of primary tissues, to regenerate tissues and in
CC chemotaxis. The polypeptides can also be used as a food additive or
CC preservative to increase or decrease storage capabilities, fat content,
CC lipid, protein, carbohydrate, vitamins, minerals, cofactors and other
CC nutritional components. AAF22307 to AAF22315 and AAB63048 represent
CC sequences used in the exemplification of the present invention
XX Sequence 184 AA;
SQ Query Match 67.0%; Score 864; DB 3; Length 184;
Best Local Similarity 98.8%; Pred. No. 1e-81;
Matches 163; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 78 SVLRGGGLTCSYRLRQVHLHWSADHSGSEHIVDGVSYAAELHVVHNSDKYPSFVEAAH 137
DB 1 NVLRGGGLTCSYRLRQVHLHWSADHSGSEHIVDGVSYAAELHVVHNSDKYPSFVYAAH 60
QY 138 EPDGLAVLGVFLQIGEPNSLOKQITDLDTSIKKKGKQTRTFNFDLLSLLPPSWDYTYPG 197
DB 61 EPDGLAVLGVFLQIGEPNSLOKQITDLDTSIKKKGKQTRTFNFDLLSLLPPSWDYTYPG 120
QY 198 SLTVPPLESVTWTVLKPINISSQQLAKFRSLCTAEGEAAFL 242
DB 121 SLTVPPLESVTWTVLKPINISSQQLAKFRSLCTAEGEAAFL 165

RESULT 7
AAO15236
ID AAO15236 standard; protein; 261 AA.
XX AAO15236;
XX AC AAO15236;
XX DT 05-SEP-2002 (first entry)
XX DE Human carbonic anhydrase I (Clnl15) protein.
XX KW Human; gastrointestinal cancer; stomach cancer; small intestine cancer;
KW colon cancer; gastrointestinal specific gene; GSG; galectin-4; Clnl14;
KW carbonic anhydrase I; Clnl15; gastrointestinal cancer marker.
XX OS Homo sapiens.
XX PN US2002042088-A1.
XX PD 11-APR-2002.
XX PF 09-MAR-2001; 2001US-00802674.
XX PR 09-MAR-2000; 2000US-0188061P.
XX PA (MACI/) MACINA R A.
PA (PIDE/) PIDERIT A.
PA (SUNY/) SUN Y.
XX Macina RA, Piderit A, Sun Y;
XX WPI; 2002-507213/54.
XX N-PSDB; AAL43637.
XX Diagnosing, monitoring, staging, imaging and treating cancers, e.g.
PT gastrointestinal cancers such as stomach, small intestine and colon
PT cancer, associated with the expression of gastrointestinal specific genes
PT Clnl14 and Clnl15.
XX Claim 7; Page 19; 23pp; English.
XX The invention comprises a method for diagnosing the presence of
CC gastrointestinal cancers (e.g. cancers of the stomach, small intestine
CC and colon) associated with two gastrointestinal specific genes (GSGs).
CC The two GSGs are human galectin-4 (Clnl14) and human carbonic anhydrase I
CC (Clnl15). It has been found that Clnl14 and Clnl15 serve as useful
CC markers in the diagnosis of gastrointestinal cancer. The method of the
CC invention is useful for detecting, diagnosing, monitoring, staging,
CC prognosticating, imaging and treating gastrointestinal cancers associated
CC with the expression of GSGs Clnl14 and Clnl15. The present amino acid
CC sequence represents the human carbonic anhydrase I (Clnl15) protein
XX Sequence 261 AA;
SQ Query Match 62.4%; Score 804; DB 5; Length 261;
Best Local Similarity 59.4%; Pred. No. 3.3e-75;
Matches 142; Conservative 44; Mismatches 53; Indels 0; Gaps 0;
QY 1 MSRLSWGYREHNGPIHWKEFFFPADGQQSQPIIKYKVKYDSLSPLSKYDPSSAKII 60
DB 1 MASPDNGYDDKNGPEQMSKLYPIANGNNSPVDIKTSETKHDTSLKPISVSYNPATAKEI 60
QY 61 SNSGHFNVDFTDTEKNKSVLRGGGLTCSYRLRQVHLHWSADHSGSEHIVDGVSYAAELH 120
DB 61 INVGHFHVNFENDNRNRSVLKGGFFSDSYKLFQFHFHWGSTNEHSGEHTVDGVKYSALH 120
QY 121 VVHNSDKYPSFVEAAHEPDGLAVLGVFLQIGEPNSLOKQITDLDTSIKKKGKQTRTFNF 180
DB 121 VAHNSAKYSLSLAESAASKADGLAVIGVLMKVGANPKLQKVLDAQAIAIKTKGRAPTFNF 180
QY 181 DLLSLLPPSWDYTYPGSLTVPPLESVTWTVLKPINISSQQLAKFRSLCTAEGEAA 239
DB 181 DPSTLLPSSLDFTWYTPGSLTHPFLYBSVTWIIKESISVSSEQLAQFRSLLSNVEGONA 239

```
RESULT 8
ADA10962
ID ADA10962 standard; protein; 261 AA.
XX
XX AC ADA10962;
XX
XX DT 06-NOV-2003 (first entry)
XX
XX DE Human cDNA differentially expressed in colon cancer #54 product.
XX
XX KW differential expression; colon cancer; cancer; human.
XX
XX OS Homo sapiens.
XX
XX PN US2002160382-A1.
XX
XX PD 31-OCT-2002.
XX
XX PF 11-OCT-2001; 2001US-00981353.
XX
XX PR 11-OCT-2000; 2000US-0239841P.
XX
XX PA (LASEK A W.
XX (JONES D A.
XX
XX PI Lasek AW, Jones DA;
XX
XX DR WPI; 2003-265756/26.
XX N-PSDB; ADA10961.
XX
XX PT New combination comprising cDNAs that are differentially expressed in
XX colon disorder, useful for diagnosing, treating, staging or monitoring
XX treatment for colon cancers.
XX
XX PS Example 14; SEQ ID NO 80; 231pp; English.
XX
XX CC The invention relates to a combination comprising cDNAs that are
XX differentially expressed in colon disorder. The methods and compositions
XX of the present invention are useful for diagnosing, treating, staging or
XX monitoring treatment for colon cancer. They are also useful in high
XX throughput methods for using cDNAs to detect differential expression of
XX nucleic acids in a sample, screening molecules or compounds to identify a
XX ligand which specifically binds a cDNA and using a protein to screen
XX molecules or compounds to identify at least one ligand which specifically
XX binds the protein. The present sequence represents the amino acid
XX sequence of a human cDNA differentially expressed in colon cancer
XX protein.
XX
XX SQ Sequence 261 AA;
Query Match 62.4%; Score 804; DB 6; Length 261;
Best Local Similarity 59.4%; Pred. No. 3.3e-75;
Matches 142; Conservative 44; Mismatches 53; Indels 0; Gaps 0;
QY 1 MSRLSWGVRHNGPHTHWKEFFPIADGQSQSPTEIKTEVKYDSSLRPLSIKYDPSAKII 60
Db 1 MASPDWGYDDKNGPEQWGLKLYPIANGNNQSPVDIKTSETKHDTSLKPISSVSNPATAKEI 60
QY 61 SNSGHSFNVDPDDTENKSVLRGGPLTGSYRLRQVHLHWGSADHGSEHIVDGVSYAEHL 120
Db 61 INVGHSHFNVFEDNDRSVLKGPPFSDSYRLRQVHLHWGSTNEHSGSEHTVDGVKYSAEHL 120
QY 121 VVHNSDKYPSFVEAAHEPDGLAVLGVLQIGEPNSQLQKIDTDLDSIKEKGKQRTFNF 180
Db 121 VAHNSAKYSSLAESAASADGLAVLGVLKMGVEANPKQLKVLDAQAIAKTKGRAPFTNF 180
QY 181 DLLSLLPSWDVWTYVPGSLTVPPLLESVTWIVLKQPINISSQQLAKFRSLICTAEGEAA 239
Db 181 DPSTLLPSLDFWTYVPGSLTVPPLLESVTWIVLKQPINISSQQLAKFRSLICTAEGEAA 239
```

```
RESULT 9
AAG73863
ID AAG73863 standard; protein; 263 AA.
XX
XX AC AAG73863;
XX
XX DT 03-SEP-2001 (first entry)
XX
XX DE Human colon cancer antigen protein SEQ ID NO:4627.
XX
XX KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX colorectal carcinoma; chromosome 8.
XX
XX OS Homo sapiens.
XX
XX PN WO200122920-A2.
XX
XX PD 05-APR-2001.
XX
XX PF 28-SEP-2000; 2000WO-US026524.
XX
XX PR 29-SEP-1999; 99US-0157137P.
XX 03-NOV-1999; 99US-0163280P.
XX
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
XX DR WPI; 2001-235357/24.
XX N-PSDB; AAG73864.
XX
XX PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers.
XX
XX PS Claim 11; Page 6427-6429; 9803pp; English.
XX
XX CC AAG732943 to AAG73514 and AAG77788 represent human colon
XX cancer-associated nucleic acid molecules (N) and proteins (P), where the
XX proteins are collectively known as colon cancer antigens. The colon
XX cancer antigens have cytostatic activity and can be used in gene therapy
XX and vaccine production. N and P may be used in the prevention, diagnosis
XX and treatment of diseases associated with inappropriate P expression. For
XX example, N and P may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of P by expressing inactive proteins or to
XX supplement the patient's own production of P. Additionally, N may be used
XX to produce the colon cancer-associated P, by inserting the nucleic acids
XX into a host cell and culturing the cell to express the proteins. N and P
XX can be used in the prevention, diagnosis and treatment of colorectal
XX carcinomas and cancers. AAG737196 to AAG737204 and AAG77789 represent
XX sequences used in the exemplification of the present invention. N.B.
XX Pages 666 to 682 and page 7053 of the sequence listing were missing at
XX time of publication, meaning no sequences are present for SEQ ID NO:1027
XX to 1052, 7921 and 7922
XX
XX SQ Sequence 263 AA;
Query Match 62.4%; Score 804; DB 4; Length 263;
Best Local Similarity 59.4%; Pred. No. 3.3e-75;
Matches 142; Conservative 44; Mismatches 53; Indels 0; Gaps 0;
QY 1 MSRLSWGVRHNGPHTHWKEFFPIADGQSQSPTEIKTEVKYDSSLRPLSIKYDPSAKII 60
Db 3 MASPDWGYDDKNGPEQWGLKLYPIANGNNQSPVDIKTSETKHDTSLKPISSVSNPATAKEI 62
QY 61 SNSGHSFNVDPDDTENKSVLRGGPLTGSYRLRQVHLHWGSADHGSEHIVDGVSYAEHL 120
Db 63 INVGHSHFNVFEDNDRSVLKGPPFSDSYRLRQVHLHWGSTNEHSGSEHTVDGVKYSAEHL 122
QY 121 VVHNSDKYPSFVEAAHEPDGLAVLGVLQIGEPNSQLQKIDTDLDSIKEKGKQRTFNF 180
Db 123 VAHNSAKYSSLAESAASADGLAVLGVLKMGVEANPKQLKVLDAQAIAKTKGRAPFTNF 182
```



```

Db      125 T-KYDGFKAQVDPGLAVLGIFLKVGSAPGLQKQVVDVLDSTKTKGSADETFNDFPRL 183
QY      186 LPPSDYWTYPGSLTPPPLLESVTWLVKQPINISSQOLAKFRSLCTAEGE 237
Db      184 LPESLDYWTYPGSLTPPPLLECWTWLVKPEISVSSEQVLKFRKLNFNGE 235

RESULT 12
ID      AAB59589
XX      AAB59589 standard; protein; 259 AA.
AC      AAB59589;
XX      AAB59589;
DT      28-MAR-2001 (first entry)
XX      Human carbonic anhydrase isoform #2.
XX      Human; protein tyrosine phosphatase; PTPase; vulnery; cytostatic;
KW      antinflammatory; antidiabetic; viral infection; inflammation; cancer;
KW      receptor-type protein tyrosine phosphatase beta; RPTPbeta;
KW      carbonic anhydrase; CAH; diabetes mellitus.
XX      Homo sapiens.
OS      US6160090-A.
PN      US6160090-A.
XX      12-DEC-2000.
PD      23-JUN-1993; 93US-00081929.
XX      11-JUL-1990; 90US-00551270.
PR      26-FEB-1991; 91US-00654188.
PR      15-OCT-1992; 92US-00961235.
PR      10-FEB-1993; 93US-00015973.
XX      (UYNV ) UNIV NEW YORK STATE.
XX      Margolis RU, Grumet MH, Barnea G, Schlessinger J;
PI      WPI; 2001-070117/08.
XX      Novel receptor type protein tyrosine phosphatase beta protein used to
PT      modulate normal cellular processes of differentiation, metabolism, cell
PT      cycle by competing with endogenous transmembrane receptors for ligands.
PS      Example; Fig 7; 45pp; English.
XX      The present sequence is given in a specification relating to a novel
CC      receptor-type protein tyrosine phosphatase beta (RTPbeta) protein or
CC      glycoprotein. The receptor and its ligands are useful for developing
CC      compounds and strategies for modifying cellular processes e.g. normal
CC      cellular processes such as differentiation, metabolism, cell cycle
CC      control, wound healing and neuronal function, cellular behaviour such as
CC      motility, migration, and contact inhibition, in addition to abnormal or
CC      potentially deleterious processes such as virus-receptor interactions,
CC      inflammation, cellular transformation to a cancerous state, and the
CC      development of Type 2, insulin independent, diabetes mellitus, under the
CC      control of the receptor protein tyrosine phosphatases. The receptors or
CC      their ligands may be used directly to modulate processes such as those
CC      mentioned above. They act to compete with endogenous transmembrane
CC      receptor molecules for available ligands, thus reducing or inhibiting
CC      ligand binding to endogenous RPTPases. RPTPases and/or their ligands, may
CC      also be used to screen for additional molecules that can act to modulate
CC      the activity of cellular processes
XX      Sequence 259 AA;
SQ      Query Match 61.2%; Score 788.5; DB 4; Length 259;
      Best Local Similarity 61.6%; Pred. No. 1.3e-73;
      Matches 143; Conservative 33; Mismatches 55; Indels 1; Gaps 1;
QY      6 WGYREHNGIHWKEFFPIADGQQSPFIEKTVKVDSSLRPLSLIKYDPSSAKI1NSGH 65

```

```

Db      4 WYGYHNGHEHWHKOPPIAKGERQSPVDIDTHTAKYDPSLKLPSVSYDQATSLRLNNGH 63
QY      66 SFNVDFDDTENKSVLRGGELTGSYRLRQVHLHWGSADHDGSEHIVDGVSYAAELHVVHWN 125
Db      64 AFNVEFDDSDQKXAVLKGGLDGTYRLIQFHFWGSLDQGSSEHTVDDKKYAAELHVLHWN 123
QY      126 SDKYPSFVEAAHEPDGLAVLGVLQIGEPNSOLOKITDITLDSIKEKGKQTRFTNFULLSL 185
Db      124 T-KYDGFKAQVDPGLAVLGIFLKVGSAPGLQKQVVDVLDSTKTKGSADETFNDFPRL 182
QY      186 LPPSDYWTYPGSLTPPPLLESVTWLVKQPINISSQOLAKFRSLCTAEGE 237
Db      183 LPESLDYWTYPGSLTPPPLLECWTWLVKPEISVSSEQVLKFRKLNFNGE 234

RESULT 13
ADE62800
ID      ADE62800 standard; protein; 259 AA.
XX      ADE62800;
XX      ADE62800;
DT      29-JAN-2004 (first entry)
XX      Human Protein P00918, SEQ ID NO 8733.
DE      Human; pain; neuronal tissue; gene therapy;
KW      spinal segmental nerve injury; chronic constriction injury; CCI;
KW      spared nerve injury; SNI; Chung.
XX      Homo sapiens.
OS      WO2003016475-A2.
PN      WO2003016475-A2.
XX      27-FEB-2003.
PD      14-AUG-2002; 2002WO-US025765.
XX      14-AUG-2001; 2001US-0312147P.
PR      01-NOV-2001; 2001US-0346382P.
PR      26-NOV-2001; 2001US-0333347P.
XX      (GEHO ) GEN HOSPITAL CORP.
PA      (FARB ) BAYER AG.
XX      Woolf C, D'urso D, Befort K, Costigan M;
PI      WPI; 2003-268312/26.
XX      GENBANK; P00918.
DR      New composition comprising two or more isolated polypeptides, useful for
PT      preparing a medicament for treating pain in an animal.
XX      Claim 1; Page; 1017pp; English.
XX      The invention discloses a composition comprising two or more isolated rat
CC      or human polynucleotides or a polynucleotide which represents a fragment,
CC      derivative or allelic variation of the nucleic acid sequence. Also
CC      claimed are a vector comprising the novel polynucleotide, a host cell
CC      comprising the vector, a method for identifying a nucleotide sequence
CC      which is differentially regulated in an animal subjected to pain and a
CC      kit to perform the method, an array, a method for identifying an agent
CC      that increases or decreases the expression of the polynucleotide sequence
CC      that is differentially expressed in neuronal tissue of a first animal
CC      subjected to pain, a method for identifying a compound which regulates
CC      the expression of a polynucleotide sequence which is differentially
CC      expressed in an animal subjected to pain, a method for identifying a
CC      compound that regulates the activity of one or more of the
CC      polynucleotides, a method for producing a pharmaceutical composition, a
CC      method for identifying a compound or small molecule that regulates the
CC      activity in an animal of one or more of the polypeptides given in the
CC      specification, a method for identifying a compound useful in treating
CC      pain and a pharmaceutical composition comprising the one or more

```


CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung)), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a human protein (shown in Table 2 of
CC the specification) which is differentially expressed during pain. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 259 AA;

Query Match 61.2%; Score 788.5; DB 7; Length 259;
Best Local Similarity 61.6%; Pred. No. 1.3e-73;
Matches 143; Conservative 33; Mismatches 55; Indels 1; Gaps 1;

QY 6 WGYREHNGPIHWKEFPPIADGQOQSPLEIKTEVKYDSSRLPLSIKYDPSSAKIISNSGH 65
DB 4 WGYGKHNGEPHEHKKOPPIAKGERQSPVDIDTHTAKYDPSLPLSVSYDQATSLRLNNGH 63
QY 66 SFNVDPDDTENKSVLRGGPLTGSYRLRQVHLHWGSDHGHSEHIVDGVSYAAELHVVHWN 125
DB 64 AFNVEFDDSDQKAVLKGGLDGTYRLIQFHFHWGSLDGQGSSEHTVDKKYAAELHLVHWN 123
QY 126 SDKYPFVEAAHEPDLGLAVLGVFLQIGEPNSQLQKITDITLDSIKEKGKQTRFTNFDLLSL 185
DB 124 T-KYDGFGRKAVQOPDGLAVLGIPLKVGSAKPGLOKVVDVLDLSIKTKGSADFTNDFPRGL 182
QY 186 LPSPDWTYPGSLTVPPLLESVTWVLKOPINISSQOLAKFRSLCTAEGE 237
DB 183 LPESLDWTYPGSLTTPPLLECVTWIVLKEPISVSSEQVLKFKLNFNGEGE 234

RESULT 14
AD663735
ID ADE663735 standard; protein; 259 AA.

AC ADE63735;
DT 29-JAN-2004 (first entry)
XX Human Protein P00918, SEQ ID NO 9679.
XX Human; pain; neuronal tissue; gene therapy;
XX spinal segmental nerve injury; chronic constriction injury; CCI;
XX spared nerve injury; SNI; Chung.
OS Homo sapiens.

FN WO2003016475-A2.
XX 27-FEB-2003.
XX 14-AUG-2002; 2002WO-US025765.
XX 14-AUG-2001; 2001US-0312147P.
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-0333347P.
XX (GENO) GEN HOSPITAL CORP.
PA (FARB) BAYER AG.

XX Woolf C, D'urso D, Befort K, Costigan M;
FI WPI; 2003-268312/26.
DR GENBANK; P00918.
XX New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.

XX Claim 1; Page; 1017pp; English.
XX The invention discloses a composition comprising two or more isolated rat

CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung)), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a human protein (shown in Table 2 of
CC the specification) which is differentially expressed during pain. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 259 AA;

Query Match 61.2%; Score 788.5; DB 7; Length 259;
Best Local Similarity 61.6%; Pred. No. 1.3e-73;
Matches 143; Conservative 33; Mismatches 55; Indels 1; Gaps 1;

QY 6 WGYREHNGPIHWKEFPPIADGQOQSPLEIKTEVKYDSSRLPLSIKYDPSSAKIISNSGH 65
DB 4 WGYGKHNGEPHEHKKOPPIAKGERQSPVDIDTHTAKYDPSLPLSVSYDQATSLRLNNGH 63
QY 66 SFNVDPDDTENKSVLRGGPLTGSYRLRQVHLHWGSDHGHSEHIVDGVSYAAELHVVHWN 125
DB 64 AFNVEFDDSDQKAVLKGGLDGTYRLIQFHFHWGSLDGQGSSEHTVDKKYAAELHLVHWN 123
QY 126 SDKYPFVEAAHEPDLGLAVLGVFLQIGEPNSQLQKITDITLDSIKEKGKQTRFTNFDLLSL 185
DB 124 T-KYDGFGRKAVQOPDGLAVLGIPLKVGSAKPGLOKVVDVLDLSIKTKGSADFTNDFPRGL 182
QY 186 LPSPDWTYPGSLTVPPLLESVTWVLKOPINISSQOLAKFRSLCTAEGE 237
DB 183 LPESLDWTYPGSLTTPPLLECVTWIVLKEPISVSSEQVLKFKLNFNGEGE 234

RESULT 15

ADAL0999

ID ADAL0999 standard; protein; 260 AA.

XX AC ADAL0999;

XX 06-NOV-2003 (first entry)

XX Human cDNA differentially expressed in colon cancer #78 product.

XX differential expression; colon cancer; cancer; human.

XX Homo sapiens.

XX US2002160382-A1.

XX 31-OCT-2002.

XX 11-OCT-2001; 2001US-00981353.

XX 11-OCT-2000; 2000US-0239841P.

XX (LASE//) LASEK A W.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 14:58:44 ; Search time 23 Seconds
(without alignments)

547.869 Million cell updates/sec

Title: US-10-069-434-1

Perfect score: 1289

Sequence: 1 MSRLSWGVRHNGPIHWKEF.....QLAKFRSLCTAGGAAAF 242

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1289	100.0	262	1 CAHD HUMAN	Q8n1q1 homo sapien
2	1187	92.1	262	1 CAHD MOUSE	Q9d6n1 mus musculus
3	848	65.8	262	1 CAH1 MONDO	Q8hy33 monodelphis
4	808.5	62.7	259	1 CAH2 TRIHK	Q8uwa5 tribolodon
5	802	62.2	260	1 CAH1 HUMAN	P00915 homo sapien
6	800	62.1	260	1 CAH1 HORSE	P00917 equus caball
7	788.5	61.2	259	1 CAH2 HUMAN	P00918 homo sapien
8	788.5	61.2	259	1 CAH2 SHEEP	P00922 ovis aries
9	785.5	60.9	259	1 CAH2 BOVIN	P00921 bos taurus
10	783.5	60.8	260	1 CAH2 BRARE	Q92051 brachydanio
11	782	60.7	260	1 CAH1 MACNE	P35217 macaca neme
12	780	60.5	260	1 CAH1 MACMU	P00916 macaca mula
13	778.5	60.4	259	1 CAH2 CHICK	P07630 gallus gall
14	775	60.1	260	1 CAH1 SHEEP	P48282 ovis aries
15	774	60.0	260	1 CAH1 MOUSE	P13634 mus musculus
16	765.5	59.4	259	1 CAH3 MOUSE	P16015 mus musculus
17	761.5	59.1	259	1 CAH3 RAT	P14141 rattus norv
18	754.5	58.5	259	1 CAH3 HORSE	P07450 equus caball
19	749.5	58.1	259	1 CAH2 RABIT	P00919 oryctolagus
20	748.5	58.1	259	1 CAH2 RAT	P07139 rattus norv
21	748.5	58.1	259	1 CAH3 HUMAN	P07451 homo sapien
22	747.5	58.0	259	1 CAH2 MOUSE	P00920 mus musculus
23	717.5	55.7	235	1 CAH1 RABIT	P07452 oryctolagus
24	699	53.5	264	1 CAH7 HUMAN	P43166 homo sapien
25	653	50.7	255	1 CAH7 MOUSE	Q9exq8 mus musculus
26	634	49.2	317	1 CASB MOUSE	Q9qza0 mus musculus
27	631	49.0	317	1 CASB HUMAN	Q9y2d0 homo sapien
28	589	45.7	305	1 CAH5 HUMAN	P35218 homo sapien
29	585	45.4	299	1 CAH5 MOUSE	P23589 mus musculus
30	581.5	45.1	304	1 CAH5 RAT	P43165 rattus norv
31	453	35.1	290	1 CAH8 MOUSE	P28651 mus musculus
32	451	35.0	289	1 CAH8 HUMAN	P35219 homo sapien
33	436.5	33.9	354	1 CAHC MOUSE	Q8ci85 mus musculus

RESULT 1

ID	CAHD HUMAN	STANDARD	PRT	262 AA
AC	Q8N1Q1;			
DT	10-OCT-2003 (Rel. 42, Created)			
DT	10-OCT-2003 (Rel. 42, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Carbonic anhydrase XIII (EC 4.2.1.1) (Carbonate dehydratase XIII) (CA-XIII).			
DE	CA13.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]_TaxID=9606;			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Tongue;			
RA	Tanigami A., Fujiwara T., Shibahara T., Goto Y., Hirao M., Shimizu F.,			
RA	Wakebe H., Ono T., Hishigaki H., Watanabe T., Ozaki K., Sugiyama T.,			
RA	Irie Y., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J.,			
RA	Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,			
RA	Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,			
RA	Wagatsuma M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,			
RA	Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K.,			
RA	Masuko Y., Nagai K., Isogai T.			
RT	"NEO human cDNA sequencing project."			
RL	Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.			
RP	[2]			
RC	SEQUENCE FROM N.A.			
RP	TISSUE=Pancreas;			
RX	MEDLINE=22388257; PubMed=12477932;			
RA	Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,			
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,			
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,			
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,			
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,			
RA	Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,			
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,			
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,			
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,			
RA	Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,			
RA	Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,			
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,			
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,			
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,			
RA	Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,			
RA	Schneerch A., Schein J.E., Jones S.J.M., Marra M.A.			
RT	"Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences."			
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).			
CC	-!- FUNCTION: Reversible hydration of carbon dioxide (By similarity).			
CC	-!- CATALYTIC ACTIVITY: H(2)CO(3) = CO(2) + H(2)O.			
CC	-!- COFACTOR: Zinc (By similarity).			
CC	-!- SIMILARITY: Belongs to the eukaryotic-type carbonic anhydrase family.			

ALIGNMENTS

```
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; AK095314; BAC04528.1; --
DR EMBL; BC052602; AAH52602.1; --
DR InterPro; IPR001148; Euk_COanhd.
DR Pfam; PF00194; carb_anhydriase; 1.
DR ProDom; PD000865; Euk_COanhd; 1.
DR PROSITE; PS00162; Euk_CO2_ANHYDRASE; 1.
KW Lyase; Zinc.
FT METAL 95 95 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 97 97 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 120 120 ZINC (CATALYTIC) (BY SIMILARITY).
SQ SEQUENCE 262 AA; 29443 MW; AE677F028ED729FE CRC64;
Query Match 100.0%; Score 1289; DB 1; Length 262;
Best Local Similarity 100.0%; Pred. No. 2.3e-105;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSRLSWGYPHNGPIHWKEFFPIADGQDSPIETKVKYDSSLRPLSIKYDPSSAKII 60
DB 1 MSRLSWGYPHNGPIHWKEFFPIADGQDSPIETKVKYDSSLRPLSIKYDPSSAKII 60
QY 61 SNSGHSFNVDFTDENKSVLRGGLTGSYRLQVHLHWSADHSGSEHIVDGVSYAAELH 120
DB 61 SNSGHSFNVDFTDENKSVLRGGLTGSYRLQVHLHWSADHSGSEHIVDGVSYAAELH 120
QY 121 VVHNSDKYPSFVEAAHEPDGLAVLGVLQIGEPNSQLKITTDLDSIKEKGKQTRFTNF 180
DB 121 VVHNSDKYPSFVEAAHEPDGLAVLGVLQIGEPNSQLKITTDLDSIKEKGKQTRFTNF 180
QY 181 DLLSLPPSWDWTYPGSLTVPPLESVTWVLKQPINISSQQLAKFRSLCTAEGEAAA 240
DB 181 DLLSLPPSWDWTYPGSLTVPPLESVTWVLKQPINISSQQLAKFRSLCTAEGEAAA 240
QY 241 FL 242
DB 241 FL 242
RESULT 2
CAHD_MOUSE STANDARD; PRT; 262 AA.
AC Q9D6N1;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Carbonic anhydrase XIII (EC 4.2.1.1) (Carbonate dehydratase XIII) (CA-
DE XIII).
GN CA13 OR CA13.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CJH;
RA Hewett-Emmett D., Shimmin L.C.;
RT "Characterization and evolution of two new members of the alpha-
RT carbonic anhydrase gene family in mouse: Car13 and Car15.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Tongue;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
```

```
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustingich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -!- FUNCTION: Reversible hydration of carbon dioxide (By similarity).
CC -!- CATALYTIC ACTIVITY: H(2)CO(3) = CO(2) + H(2)O.
CC -!- COFACTOR: Zinc (By similarity).
CC -!- SIMILARITY: Belongs to the eukaryotic-type carbonic anhydrase
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; AF231123; AK16672.1; --
DR EMBL; AK010166; BAB26742.1; --
DR HSSP; P00918; 1C1M.
DR MGD; MGI:1931322; Carl3.
DR InterPro; IPR001148; Euk_COanhd.
DR Pfam; PF00194; carb_anhydriase; 1.
DR ProDom; PD000865; Euk_COanhd; 1.
DR PROSITE; PS00162; Euk_CO2_ANHYDRASE; 1.
KW Lyase; Zinc.
FT METAL 95 95 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 97 97 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 120 120 ZINC (CATALYTIC) (BY SIMILARITY).
SQ SEQUENCE 262 AA; 29522 MW; E3CA4674C1CF4A12 CRC64;
Query Match 92.1%; Score 1187; DB 1; Length 262;
Best Local Similarity 91.7%; Pred. No. 1.9e-96;
Matches 222; Conservative 8; Mismatches 12; Indels 0; Gaps 0;
QY 1 MSRLSWGYPHNGPIHWKEFFPIADGQDSPIETKVKYDSSLRPLSIKYDPSSAKII 60
DB 1 MARLSWGYGEHNGPIHWNELFPIADGQDSPIETKVKYDSSLRPLSIKYDPSSAKII 60
QY 61 SNSGHSFNVDFTDENKSVLRGGLTGSYRLQVHLHWSADHSGSEHIVDGVSYAAELH 120
DB 61 SNSGHSFNVDFTDENKSVLRGGLTGSYRLQVHLHWSADHSGSEHIVDGVSYAAELH 120
QY 121 VVHNSDKYPSFVEAAHEPDGLAVLGVLQIGEPNSQLKITTDLDSIKEKGKQTRFTNF 180
DB 121 VVHNSDKYPSFVEAAHEPDGLAVLGVLQIGEPNSQLKITTDLDSIKEKGKQTRFTNF 180
QY 181 DLLSLPPSWDWTYPGSLTVPPLESVTWVLKQPINISSQQLAKFRSLCTAEGEAAA 240
DB 181 DPICLLPSSWDWTYPGSLTVPPLESVTWVLKQPINISSQQLAKFRSLCTAEGEAAA 240
QY 241 FL 242
DB 241 FL 242
RESULT 3
CAH1_MONDO
```


RT the C isozyme.";
RL J. Biol. Chem. 249:2329-2337(1974).
RN [2]
RP SEQUENCE.
RX MEDLINE=77006079; PubMed=823150;
RA Henderson L.E., Henriksson D., Nyman P.O.;
RT "Primary structure of human carbonic anhydrase C.";
RL J. Biol. Chem. 251:5457-5463(1976).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=87231043; PubMed=3108857;
RA Montgomery J.C., Venta P.J., Tashian R.E., Hewett-Emmett D.;
RT "Nucleotide sequence of human liver carbonic anhydrase II cDNA.";
RL Nucleic Acids Res. 15:4687-4687(1987).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=88085190; PubMed=3121496;
RA Murakami H., Marelich G.P., Grubb J.H., Kyle J.W., Sly W.S.;
RT "Cloning, expression, and sequence homologies of cDNA for human carbonic anhydrase II.";
RL Genomics 1:159-166(1987).
RN [5]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Ronald M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C., Raha S., Iqbal N.A., Peters G.J., Abramson R.D., Mullenbach S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Viallon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [6]
RP SEQUENCE OF 1-76 FROM N.A.
RX MEDLINE=86077780; PubMed=3000449;
RA Venta P.J., Montgomery J.C., Hewett-Emmett D., Tashian R.E.;
RT "Comparison of the 5' regions of human and mouse carbonic anhydrase II genes and identification of possible regulatory elements.";
RL Biochim. Biophys. Acta 846:195-201(1985).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX MEDLINE=72111787; PubMed=4621826;
RA Liljas A., Kannan K.K., Bergsten P.-C., Waara I., Fridborg K., Strandberg B., Carlsson U., Jærup L., Loevgren S., Petef M.;
RT "Crystal structure of human carbonic anhydrase C.";
RL Nature New Biol. 235:131-137(1972).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX MEDLINE=89315726; PubMed=3151019;
RA Eriksson A.E., Jones T.A., Liljas A.;
RT "Refined structure of human carbonic anhydrase II at 2.0-A resolution.";
RL Proteins 4:274-282(1988).
RN [9]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX MEDLINE=89315727; PubMed=3151020;
RA Eriksson A.E., Kylsten P.M., Jones T.A., Liljas A.;
RT "Crystallographic studies of inhibitor binding sites in human carbonic anhydrase II: a pentacoordinated binding of the SCN-ion to the zinc at high pH.";
RN [10]
RP Proteins 4:283-293(1988).
RN [10]
RP X-RAY CRYSTALLOGRAPHY (2.25 ANGSTROMS).
RX MEDLINE=98200459; PubMed=9541386;
RA Stams T., Chen Y., Boriack-Sjodin P.A., Hurt J.D., Liao J., May J.A., Dean T., Laipis P., Silverman D.N., Christianson D.W.;
RT "Structures of murine carbonic anhydrase IV and human carbonic anhydrase II complexed with brinzolamide: molecular basis of isozyme-drug discrimination.";
RL Protein Sci. 7:556-563(1998).
RN [11]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX MEDLINE=20530299; PubMed=11076507;
RA Cox J.D., Hunt J.A., Compher K.M., Fierke C.A., Christianson D.W.;
RT "Structural influence of hydrophobic core residues on metal binding and specificity in carbonic anhydrase II.";
RL Biochemistry 39:13687-13694(2000).
RN [12]
RP VARIANT JOGJAKARTA.
RX MEDLINE=83100296; PubMed=6817747;
RA Jones G.L., Sofro A.S.M., Shaw D.C.;
RT "Chemical and enzymological characterization of an Indonesian variant of human erythrocyte carbonic anhydrase II, CAII Jogjakarta (17 Lys leads to Glu).";
RL Biochem. Genet. 20:979-1000(1982).
RN [13]
RP VARIANT MELBOURNE.
RX MEDLINE=83236368; PubMed=6407977;
RA Jones G.L., Shaw D.C.;
RT "A chemical and enzymological comparison of the common major human erythrocyte carbonic anhydrase II, its minor component, and a new genetic variant, CA II Melbourne (237 pro leads to His).";
RL Hum. Genet. 63:392-399(1983).
RN [14]
RP VARIANT CA2 DEFICIENCY TVR-106.
RX MEDLINE=92026087; PubMed=128091;
RA Venta P.J., Welty R.J., Johnson T.M., Sly W.S., Tashian R.E.;
RT "Carbonic anhydrase II deficiency syndrome in a Belgian family is caused by a point mutation at an invariant histidine residue (107 His-->Tyr): complete structure of the normal human CA II gene.";
RL Am. J. Hum. Genet. 49:1082-1090(1991).
RN [15]
RP VARIANT CA2 DEFICIENCY TVR-106.
RX MEDLINE=92179270; PubMed=1542674;
RA Roth D.E., Venta P.J., Tashian R.E., Sly W.S.;
RT "Molecular basis of human carbonic anhydrase II deficiency.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:1804-1808(1992).
RN [16]
RP VARIANT CA2 DEFICIENCY TVR-106.
RX MEDLINE=96431156; PubMed=8834238;
RA Soda H., Yukizane S., Yoshida I., Koga Y., Aramaki S., Kato H.;
RT "A point mutation in exon 3 (His 107-->Tyr) in two unrelated Japanese patients with carbonic anhydrase II deficiency with central nervous system involvement.";
RL Hum. Genet. 97:435-437(1996).
RN [17]
RP VARIANT CA2 DEFICIENCY PRO-91.
RX MEDLINE=97288992; PubMed=9143915;
RA Hu P.Y., Lim E.J., Ciccolella J., Strisciuglio P., Sly W.S.;
RT "Seven novel mutations in carbonic anhydrase II deficiency syndrome identified by SSCP and direct sequencing analysis.";
RL Hum. Mutat. 9:383-387(1997).
RN [18]
RP FUNCTION: Reversible hydration of carbon dioxide.
CC -1- CATALYTIC ACTIVITY: H(2)CO(3) = CO(2) + H(2)O.
CC -1- COPACTOR: Zinc.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- DISEASE: Defects in CA2 are the cause of carbonic anhydrase II deficiency (CA2 deficiency) [MIM:259730]. It can be associated with osteopetrosis, renal tubular acidosis, cerebral calcification (marble brain disease) and mental retardation. Inheritance is autosomal dominant.
CC -1- SIMILARITY: Belongs to the eukaryotic-type carbonic anhydrase family.
CC

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M77181; AAAS1909.1; -.
DR EMBL; M77176; AAAS1909.1; JOINED.
DR EMBL; M77177; AAAS1909.1; JOINED.
DR EMBL; M77178; AAAS1909.1; JOINED.
DR EMBL; M77179; AAAS1909.1; JOINED.
DR EMBL; M77180; AAAS1909.1; JOINED.
DR EMBL; Y00339; CAA68426.1; -.
DR EMBL; X03251; CAA27012.1; -.
DR EMBL; J03037; AAAS1908.1; -.
DR EMBL; BC011949; AAH11949.1; -.
DR EMBL; M36532; AAAS1911.1; -.
DR PIR; A27175; CRH02.
DR PDB; 12CA; 15-OCT-92.
DR PDB; 1A42; 23-MAR-99.
DR PDB; 1AM6; 24-JUN-98.
DR PDB; 1AVN; 24-DEC-97.
DR PDB; 1BCD; 31-OCT-93.
DR PDB; 1BIC; 31-OCT-93.
DR PDB; 1BN1; 18-MAY-99.
DR PDB; 1BN3; 18-MAY-99.
DR PDB; 1BN4; 18-MAY-99.
DR PDB; 1BNM; 18-MAY-99.
DR PDB; 1BNN; 18-MAY-99.
DR PDB; 1BNQ; 15-JUN-99.
DR PDB; 1BNU; 15-JUN-99.
DR PDB; 1BNU; 15-JUN-99.
DR PDB; 1BNV; 16-FEB-99.
DR PDB; 1BNW; 15-JUN-99.

Query Match      61.2%; Score 788.5; DB 1; Length 259;
Best Local Similarity 61.6%; Pred. No. 1.2e-61;
Matches 143; Conservative 33; Mismatches 55; Indels 1; Gaps 1;

QY 6 WGYREHNGPIHWKEFFPIADGQSPFIEIKTEVKYDSSLRPLSIKYDPSSAKIISNGH 65
Db 4 WGYGHEHNGPEHWHKDFPIADGERQSPVDITHTAKYDPSLKSLSVSDQATSLRIANNHG 63

QY 66 SFNVDFDPTENKSVLRGGPLTGSYRLRQVHLHWGSDHGSSEHIVDGVSYAAELHVVHWN 125
Db 64 AFNVEFDSQDKAVLKGGLDGTYRLIQPHFWGSLDGGSEHTVDKYYAAELHVVHWN 123

QY 126 SDKPSFVEAAHEPDGLAVLGVLQIGEPNSOLOKITDPLDSIKEKGKQTRFTNFDLLSL 185
Db 124 T-KYGFDFGKAVQPPDGLAVLGVLGFKVGSAPGLQKVDVLDLSIKTKGKSADFTNFDPRGL 182

QY 186 LPSPSDYWTYPGSLTPPPLLESVTWVLKQPINISSOQLAKFRSLCTAEGE 237
Db 183 LPESLDYWTYPGSLTPPPLLECVTWVLKEPISVSSEQVLFKPKLNFNNGE 234

RESULT 8
CAH2 SHEEP
ID CAH2 SHEEP STANDARD; PRT; 259 AA.
AC P00922;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE Carbonic anhydrase II (EC 4.2.1.1) (Carbonate dehydratase II) (CA-II).
GN CA2.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Caprinae; Ovis.
NCBI_TaxID=9940;

```

```

RN [1]
RP SEQUENCE.
RX MEDLINE=75054988; PubMed=4215456;
RA Tanis R.J., Ferrell R.E., Tashian R.E.;
RT "Amino acid sequence of sheep carbonic anhydrase C.";
RL Biochim. Biophys. Acta 371:534-548(1974).
RN [2]
RP VARIANT FORM.
RX MEDLINE=79145542; PubMed=106895;
RA Mallet B., Gulian J.M., Sciaky M., Laurent G., Charrel M.;
RT "Multiple molecular forms of erythrocyte carbonic anhydrase of
RL sheep.";
RL Biochim. Biophys. Acta 576:290-304(1979).
CC -!- FUNCTION: Reversible hydration of carbon dioxide.
CC -!- CATALYTIC ACTIVITY: H(2)CO(3) = CO(2) + H(2)O.
CC -!- COFACTOR: Zinc.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- MISCELLANEOUS: ONE MINOR AND THREE MAJOR FORMS WERE ISOLATED
CC CHROMATOGRAPHICALLY.
CC -!- SIMILARITY: Belongs to the eukaryotic-type carbonic anhydrase
CC family.
DR PIR; A01145; CRSH2.
DR HSP; P00918; 1CIM.
DR InterPro; IPR001148; Euk Coanhd.
DR Pfam; PF00194; carb_anhydrase; 1.
DR ProDom; PD000865; Euk_coanhd; 1.
DR PROSITE; PS00162; EUK_CO2_ANHYDRASE; 1.
KW Lyase; Zinc; Acetylation.
FT MOD_RES 1 ACETYLATION.
FT METAL 93 ZINC (CATALYTIC).
FT METAL 95 ZINC (CATALYTIC).
FT METAL 118 ZINC (CATALYTIC).
FT METAL 118 ZINC (CATALYTIC).
FT VARIANT 35 K -> T (IN ONE OF THE MAJOR FORMS).
SQ SEQUENCE 259 AA; 29080 MW; E0B1DD6E67263604 CRC64;

Query Match      61.2%; Score 788.5; DB 1; Length 259;
Best Local Similarity 62.1%; Pred. No. 1.2e-61;
Matches 144; Conservative 39; Mismatches 46; Indels 1; Gaps 1;

QY 6 WGYREHNGPIHWKEFFPIADGQSPFIEIKTEVKYDSSLRPLSIKYDPSSAKIISNGH 65
Db 4 WGYGHEHNGPEHWHKDFPIADGERQSPVDITHTAKYDPSLKSLSVSDQATSLRIANNHG 63

QY 66 SFNVDFDPTENKSVLRGGPLTGSYRLRQVHLHWGSDHGSSEHIVDGVSYAAELHVVHWN 125
Db 64 SFNVEFDSQDKAVLKGGLTGTYRLVQPHFWGSSDDQSGSEHTVDRKKYAAELHVVHWN 123

QY 126 SDKPSFVEAAHEPDGLAVLGVLQIGEPNSOLOKITDPLDSIKEKGKQTRFTNFDLLSL 185
Db 124 T-KYGFDFGKAVQPPDGLAVLGVLGFKVGSAPGLQKVDVLDLSIKTKGKSADFTNFDPRGL 182

QY 186 LPSPSDYWTYPGSLTPPPLLESVTWVLKQPINISSOQLAKFRSLCTAEGE 237
Db 183 LKRALNYWTYPGSLTPPPLLESVTWVLKEPISVSSEQVLFKPKLNFNNGE 234

RESULT 9
CAH2 BOVIN
ID CAH2 BOVIN STANDARD; PRT; 259 AA.
AC P00921;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE Carbonic anhydrase II (EC 4.2.1.1) (Carbonate dehydratase II) (CA-II).
GN CA2.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
NCBI_TaxID=9913;
RN [1]
RP SEQUENCE.
RC TISSUE=Erythrocyte;

```


[illegible]

us-10-069-434-1.rsp

Tue Sep 14 16:50:57 2004

Db 185 LPSSLDYWTYFGSLTHPPLEHESVTWICKDSISPEQLAQIRGLLSAEGESA 238

Search completed: September 9, 2004, 15:11:56
Job time : 29 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 15:06:00 ; Search time 115 Seconds
(without alignments)
663.960 Million cell updates/sec

Title: US-10-069-434-1

Perfect score: 1289

Sequence: 1 MSRLSWGVRHNGPIHWKEP.....QLAKFRSLCTAGGEAAAPL 242

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:*

1: sp archaea:*

2: sp bacteria:*

3: sp fungi:*

4: sp human:*

5: sp invertebrate:*

6: sp mammal:*

7: sp mhc:*

8: sp organelle:*

9: sp phage:*

10: sp plant:*

11: sp rodent:*

12: sp virus:*

13: sp vertebrate:*

14: sp unclassified:*

15: sp virus:*

16: sp bacteriap:*

17: sp archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	800	62.1	261	13	Q8JG56	Q8JG56 lepisosteus
2	794.5	61.6	260	6	Q865Y7	Q865Y7 bos taurus
3	775.5	60.2	260	13	Q7T2K6	Q7T2K6 oncorhynchus
4	751.5	58.3	260	11	Q7TP21	Q7TP21 mus musculus
5	745.5	57.8	260	13	Q8AVG8	Q8AVG8 xenopus lae
6	721.5	56.0	260	13	Q7ZU06	Q7ZU06 xenopus lae
7	675	52.4	264	11	Q811X4	Q811X4 mus musculus
8	666	51.7	306	13	Q7ZUE2	Q7ZUE2 brachydanio
9	551	42.7	208	4	Q86YU0	Q86YU0 homo sapien
10	548.5	42.6	261	5	Q9XZG6	Q9XZG6 anthopleura
11	547	42.4	270	5	Q9V396	Q9V396 drosophila
12	546.5	42.4	192	11	Q9DCT3	Q9DCT3 mus musculus
13	487.5	37.8	243	5	Q8MPH8	Q8MPH8 riftia pach
14	436.5	33.9	344	11	Q8K2J1	Q8K2J1 mus musculus
15	406	31.5	259	13	Q93587	Q93587 platichthys
16	393	30.5	304	12	Q66218	Q66218 cowpox viru

17	392	30.4	304	12	Q8QMV3	Q8QMV3 cowpox viru
18	392	30.4	304	12	Q9JFAL	Q9JFAL vaccinia vi
19	389	30.2	304	12	Q90197	Q90197 monkeypox v
20	388	30.1	304	12	Q80DX3	Q80DX3 cowpox viru
21	386	29.9	304	12	Q8JLB8	Q8JLB8 ectromelia
22	385	29.9	304	12	Q8V4Y0	Q8V4Y0 monkeypox v
23	385	29.9	304	12	Q8V2R1	Q8V2R1 camelpox vi
24	385	29.9	304	12	Q66281	Q66281 camelpox vi
25	381	29.6	304	12	O57211	O57211 vaccinia vi
26	379	29.4	327	5	Q9W316	Q9W316 drosophila
27	376.5	29.2	110	13	Q7SX08	Q7SX08 pseudopleur
28	375	29.1	303	12	Q83439	Q83439 ectromelia
29	374.5	28.1	285	12	Q9Q8Z4	Q9Q8Z4 Shope fibro
30	365.5	28.4	320	6	Q865C0	Q865C0 canis famil
31	364.5	28.3	286	12	Q9Q8L7	Q9Q8L7 myxoma viru
32	354	27.5	317	11	Q7TNG9	Q7TNG9 mus musculu
33	354	27.5	325	11	Q80YB7	Q80YB7 mus musculu
34	349	27.1	335	5	Q9VTU8	Q9VTU8 drosophila
35	339.5	26.3	328	6	Q9N085	Q9N085 macaca fasc
36	338	26.2	250	5	Q9W3C8	Q9W3C8 drosophila
37	337.5	26.2	252	16	Q9CJT6	Q9CJT6 pasteurella
38	323.5	25.1	311	5	Q9VB76	Q9VB76 drosophila
39	315.5	24.5	328	11	Q811X3	Q811X3 rattus norv
40	312	24.2	312	13	Q7SYW3	Q7SYW3 xenopus lae
41	306.5	23.8	328	6	Q866X7	Q866X7 bos taurus
42	306	23.7	331	6	Q866X6	Q866X6 sus scrofa
43	303	23.5	1576	13	Q9I908	Q9I908 xenopus lae
44	303	23.5	2271	13	Q9I909	Q9I909 xenopus lae
45	300.5	23.3	275	16	Q9KFW1	Q9KFW1 bacillus ha

ALIGNMENTS

RESULT 1

Q8JG56	PRELIMINARY;	PRT;	261 AA.
ID	Q8JG56		
AC	Q8JG56;		
DT	01-OCT-2002 (TREMREL. 22, Created)		
DT	01-OCT-2002 (TREMREL. 22, Last sequence update)		
DT	01-OCT-2003 (TREMREL. 25, Last annotation update)		
DE	Erythrocyte carbonic anhydrase.		
OS	Lepisosteus osseus (long-nosed gar).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Actinopterygii; Neopterygii; Semionotiformes; Lepisosteidae;		
OC	Lepisosteus.		
OX	NCBI_TaxID=34771;		
RN	(1)		
RP	SEQUENCE FROM N.A.		
RA	Lund S.G., Dymont P., Gervais M.R., Moyes C.D., Tufts B.L.;		
RT	"Characterization of erythrocyte carbonic anhydrase in an ancient fish, the longnose gar (Lepisosteus osseus).";		
RL	J. Comp. Physiol. B, Biochem. Syst. Environ. Physiol. 0:0-0(2002).		
DR	EMBL; AY125007; AAM94169.1;		
DR	GO; GO:0004089; F:carbonate dehydratase activity; IEA.		
DR	GO; GO:0008270; F:zinc ion binding; IEA.		
DR	GO; GO:0006730; P:one-carbon compound metabolism; IEA.		
DR	InterPro; IPR001148; Euk Coanhd.		
DR	Pfam; PF00194; carb anhydrase; 1.		
DR	ProDom; PD000865; Euk Coanhd; 1.		
DR	PROSITE; PS00162; Euk_CO2_ANHYDRASE; 1.		
SQ	SEQUENCE 261 AA; 28700 MW; 55BDBF4A08D9F54B CRC64;		

Query Match 62.1%; Score 800; DB 13; Length 261;

Best Local Similarity 63.4%; Pred. No. 2.7e-63;

Matches 149; Conservative 30; Mismatches 56; Indels 0; Gaps 0;

Qy 5 SWGYRENGPIHWKEFFPIADGQSQSPIDIVPSQAQHPDLKPLRIYDPTSGILNNG 64

Db 4 SWGYAANGPDKWKEKFPKQSPIDIVPSQAQHPDLKPLRIYDPTSGILNNG 63

Qy 65 HSFNVDPTDTEKSVLGGPLTGSYRURVHLHWGSADHGSEHIVDGVSAAEHVVHW 124

Db	64	HSFQVDFADENDSSTLQGGPISGVYRLRQGFHFWGASDERGSEHTVGGVKVAAELHLVHW	123
QY	125	NSDKYPSFVEAAHPDGLAVLGVLQIGEPNSQLKIDTDLDSIKEKGKQTRFTNFDLLS	184
Db	124	NAGKYASFGDAKAPDGLAVGVFLKIGASPENLQKVLDAIKTKGQTPFPQDFPKI	183
QY	185	LLPPSWDYWTYPGSLTPPLLESVTWIVLKQPINISSQOLAKFRSLLCTASGEAA	239
Db	184	LLPSSLDFWTYEGSLTTPPLESVTWIVLKPEITVSSEQMAKFRSLLFTASGETA	238
RESULT 2			
ID	Q865Y7	PRELIMINARY;	PRT; 260 AA.
AC	Q865Y7;		
DT	01-JUN-2003	(TrEMBLrel. 24, Created)	
DT	01-JUN-2003	(TrEMBLrel. 24, Last sequence update)	
DT	01-OCT-2003	(TrEMBLrel. 25, Last annotation update)	
DE	Carbonic anhydrase II (EC 4.2.1.1).		
OS	Bos taurus (Bovine).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;		
OC	Bovidae; Bovinae; Bos.		
OX	NCBI_TaxID=9913;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Bone marrow;		
RA	Daigle R., Castro I., Desrochers M., Charest P.-M.;		
RT	"Full length cDNA of Bovine Carbonic Anhydrase II.",		
RL	Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AY240020; AAC85140.1; -		
DR	GO; GO:004089; F:carbonate dehydratase activity; IEA.		
DR	GO; GO:0016829; F:lyase activity; IEA.		
DR	GO; GO:0008270; F:zinc ion binding; IEA.		
DR	GO; GO:0006730; P:one-carbon compound metabolism; IEA.		
DR	InterPro; IPR001148; Euk_Coanhd.		
DR	Pfam; PF00194; carb anhydrase; 1.		
DR	ProDom; PD000865; Euk_Coanhd; 1.		
DR	PROSITE; PS00162; EUK_CO2_ANHYDRASE; 1.		
KW	Lyase.		
SQ	SEQUENCE	260 AA; 29114 MW; A89412C23FFD19A7	CRC64;
Query Match 61.6%; Score 794.5; DB 6; Length 260;			
Best Local Similarity 62.5%; Pred. No. 8.2e-63;			
Matches 145; Conservative 39; Mismatches 47; Indels 1; Gaps 1;			
QY	6	WGVRHNGPIHWKEFFPIADGQSPTEIKTKVKYDSSLRPLSIKYDPSSAKIISNSGH	65
Db	5	WGYGKNGPDPHWKDPFIANGERSQSPVIDTKAVQDPALKPLALVYGEATSRMWNHGH	64
QY	66	SFNVDFTDNTENKSVLRGGPITGSYRLRQVHLHWGSDDHGSEHIVDGVSYAAELHVVHN	125
Db	65	SFNVEYDSDQKAVLKDGPITGYRLVQPFHFWGSSDQGSHTVDRKKYAAELHVVHN	124
QY	126	SDKYPSFVEAAHPDGLAVLGVLQIGEPNSQLKIDTDLDSIKEKGKQTRFTNFDLLSL	185
Db	125	T-KYGFEGTAAQPDGLAVGVFLKIGASPENLQKVLDAIKTKGQTPFPQDFPKI	183
QY	186	LPSPSWDYWTYPGSLTPPLLESVTWIVLKQPINISSQOLAKFRSLLCTARGE	237
Db	184	LPNVLDFTWTYPGSLTTPPLESVTWIVLKPEITVSSEQMAKFRILFNARGE	235
RESULT 3			
ID	Q7T2K6	PRELIMINARY;	PRT; 260 AA.
AC	Q7T2K6;		
DT	01-OCT-2003	(TrEMBLrel. 25, Created)	
DT	01-OCT-2003	(TrEMBLrel. 25, Last sequence update)	
DT	01-OCT-2003	(TrEMBLrel. 25, Last annotation update)	
DE	Erythrocyte carbonic anhydrase.		
OS	Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		

[illegible]

DR Pfam; PF00194; carb anhydrase; 1.
DR ProDom; PD000865; Euk_COanhnd; 1.
DR PROSITE; PS00162; EUK_CO2 ANHYDRASE; 1.
SQ SEQUENCE 264 AA; 29915 MW; B58E0E20CB840FA5 CRC64;

Query Match 52.4%; Score 675; DB 11; Length 264;
Best Local Similarity 52.2%; Pred. No. 4e-52;
Matches 121; Conservative 47; Mismatches 64; Indels 0; Gaps 0;

QY 6 WGYREHNGPIHWKEFFFIADGDOOSPIETKTEVKYDSSLRPLSIKYDPSSAKIISNGH 65
Db |||:||||:|:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
QY 7 WGYQDDGPNWHKLYPIAGDQSPINIISSQAVYSPSLQPLELYEACMSLSITNGH 66
Db |||:||||:|:||||:||||:||||:||||:||||:||||:||||:||||:|
QY 66 SFNVDFDPTENKSVLRGGPITGYSRLRQVHLHWGSDHGGSEHIVDGVSYAELHVVHWN 125
Db |||:||||:|:||||:||||:||||:||||:||||:||||:||||:||||:|
QY 67 SVQVDFDSDRTVVGSGPLEPGYRLKQLHFHGGKRDGMSEHTVDPKSPSELHLVHWN 126
Db |||:||||:|:||||:||||:||||:||||:||||:||||:||||:||||:|

QY 126 SDKYPSEFAAHEPDGLAVLGVLQIGEPNSQLKQITDLDLSIKEKGKQTRFTNFDLLSL 185
Db |||:||||:|:||||:||||:||||:||||:||||:||||:||||:||||:|

QY 127 AKKSTTGEAAAPDGLAVGVFLETGDEHPSNMRLTDALYVWRFKDTKAQFSCFPKCL 186
Db |||:||||:|:||||:||||:||||:||||:||||:||||:||||:||||:|

QY 186 LPSPSWDYWTYPGSLTVPPLLESVTWIVLKQPINISSQOLAKFRSLICTAEGE 237
Db |||:||||:|:||||:||||:||||:||||:||||:||||:||||:||||:|

QY 187 LPTSRHYWTYPGSLTTPPLSESVTWIVLRPEIRISERQMBKFRSLFTSEDD 238
Db |||:||||:|:||||:||||:||||:||||:||||:||||:||||:||||:|

RESULT 8
QY Q7ZUE2 PRELIMINARY; PRT; 306 AA.
AC Q7ZUE2;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to carbonic anhydrase VII (fragment).
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN 1;
RP SEQUENCE FROM N.A.
RC TISSUE=Body;
RA Strausberg R.;
RL EMBL; BC049309; AAH49309.1; -.
DR GO; GO:0004089; F:carbonate dehydratase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006730; P:one-carbon compound metabolism; IEA.
DR InterPro; IPR001148; Euk_COanhnd.
DR Pfam; PF00194; carb anhydrase; 1.
DR ProDom; PD000865; Euk_COanhnd; 1.
DR PROSITE; PS00162; EUK_CO2 ANHYDRASE; 1.
FT NON_TER 1;
SQ SEQUENCE 306 AA; 34587 MW; B35078B06365C0B CRC64;

Query Match 51.7%; Score 666; DB 13; Length 306;
Best Local Similarity 53.0%; Pred. No. 3.2e-51;
Matches 123; Conservative 40; Mismatches 69; Indels 0; Gaps 0;

QY 6 WGYREHNGPIHWKEFFFIADGDOOSPIETKTEVKYDSSLRPLSIKYDPSSAKIISNGH 65
Db |||:||||:|:||||:||||:||||:||||:||||:||||:||||:||||:|
QY 49 WGYGDNGPSAHHKDYPTAGNRQSPIDIVPSEAVFDKLSPIALSYNNCTSLISNGH 108
Db |||:||||:|:||||:||||:||||:||||:||||:||||:||||:||||:|

QY 66 SFNVDFDPTENKSVLRGGPITGYSRLRQVHLHWGSDHGGSEHIVDGVSYAELHVVHWN 125
Db |||:||||:|:||||:||||:||||:||||:||||:||||:||||:||||:|

QY 109 SVVFEVFDTERSVITGGPLENMYLKLQPHFWGSKGCGSEHTVAGTKTFVSELHLVHWN 168
Db |||:||||:|:||||:||||:||||:||||:||||:||||:||||:||||:|

QY 126 SDKYPSEFAAHEPDGLAVLGVLQIGEPNSQLKQITDLDLSIKEKGKQTRFTNFDLLSL 185
Db |||:||||:|:||||:||||:||||:||||:||||:||||:||||:||||:|

QY 169 ANKYKFSFAAVAPDGLAVLGLETGDEHRLHQITDALYVWRFKGLSABFKGNPKCL 228
Db |||:||||:|:||||:||||:||||:||||:||||:||||:||||:||||:|

QY 186 LPSPSWDYWTYPGSLTVPPLLESVTWIVLKQPINISSQOLAKFRSLICTAEGE 237
Db |||:||||:|:||||:||||:||||:||||:||||:||||:||||:||||:|

Db 229 LPNSLEYWTYPGSLTTPPLSESVTWIVLKQPINISSQOLAKFRSLFTSEDD 280

```

DR EMBL; AF140537; AAD32675.1; -.
DR HSSP; P00918; 1BV3.
DR GO; GO:0004089; F:carbonate dehydratase activity; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006730; P:one-carbon compound metabolism; IEA.
DR InterPro; IPR001148; Euk_Coanhd.
DR Pfam; PF00194; carb. anhydrase; 1.
DR ProDom; PD000865; Euk_Coanhd; 1.
DR PROSITE; PS00162; EUK_CO2_ANHYDRASE; 1.
KW Lyase; Zinc.
SQ SEQUENCE 261 AA; 28612 MW; 9CB4C82065DD9740 CRC64;

Query Match
Best Local Similarity 42.6%; Score 548.5; DB 5; Length 261;
Matches 109; Conservative 43; Mismatches 77; Indels 5; Gaps 4;

QY 1 MSRLSWGVRHNGPIHWKEFFFIADQDQSPTEIKTEVKYDSSL--RPLSIKIDPSSAK 58
DQ :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1 MAAPKVGYPNGGPKWAKDFPAAGARQSPIDIKTHDAQHSALKIKPLKIYSGQND 60
QY 59 IISNGHGFNVDFDTEKNKSVLRGGPLTGSYRLRQVHLHWSADHSGSEHIVDGSYAAE 118
DQ :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 61 NVTNNGYSLIVSRKTSEGTN-LSCGPLEHNYRFEQPHFWGKTSGSGSEHLLDGKAPPAE 119
QY 119 LHVHWNDSKDYSEFVEAAHEPGLAVLGVFLGIGPNSLOLKITDITLDSIKKKG-KQTRF 177
DQ :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 120 LHLVHNWNTDLFSFGGEAASSKNGLVAGVAFVIGGESAGLKTITDILPQVQVNIQDKDLK 179
QY 178 TNFDLLSLPPSW-DYWTYPGSLTPVPLESVTWLVKOPINISSQOLAKPRSL 230
DQ :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 180 VPFNLSLLPSNTNDYWTYSGSLTPPCYVESVWFKEPIHATENQMQPRSL 233

RESULT 11
ID Q9V396 PRELIMINARY; PRT; 270 AA.
AC Q9V396;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CAH1 protein (EC 4.2.1.1) (Carbonic anhydrase) (Carbonate
DE dehydratase).
GN CAH1 OR BG:DS00941.1 OR CG7820.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abriil J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA de Pablo S., Dawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Chablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Durbin K.J., Evans M.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Foster C., Gabelian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

```

Query Match 42.4%; Score 547; DB 5; Length 270;

Best Local Similarity 49.3%; Pred. No. 1.2e-40;		Matches 113; Conservative 24; Mismatches 86; Indels 6; Gaps 4;	
QY	6 WGYREHNGPIHWKEFFPIADGQDQGPPIETKKEVYDSSLR--PLSIKYDPSSAKIISNS 63		
DB	5 WGYTEENGPAHWAKYEPQASGHRQSPVITPSSAKGSELNVAPLKWKYVPEHTKSLNVP 64		
QY	64 GHSFNVDFTDENKSVLRGGPVTGS-YRLRQVHLHWSADHGSBHIIVDGVSYAAELHVV 122		
DB	65 GYCMRVVDNAGD--SELTGGPLGDQIFKLEQPHCHWGCTDSKGSBHTVDGVSYSGELHLV 122		
QY	123 HNSDKYSFVEAAHEPDGLAVLGVFLQIGEPNSQLQKTDITDLSIKEKG-KQTRFTNFD 181		
DB	123 HNWTTTKYSFGFAAAPDGLAVLGVFLKAGNHAELEDKVTLSLQVFLHKGDRVTLPQCGD 182		
QY	182 LLSLLPSPSDYVWYTPGSLTVPPLLESVTWIVLKPQINISSQQLAKFRSL 230		
DB	183 PGQLLPDVHTYTYEGSLTTPCSESIVIVFKTPIEVSDDDQLNAMENL 231		
RESULT 12			
ID	Q9DCT3 PRELIMINARY; PRT; 192 AA.		
AC	Q9DCT3;		
DT	01-JUN-2001 (TrEMBLrel. 17, Created)		
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)		
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)		
DE	Adult male kidney cDNA, RIKEN full-length enriched library,		
DE	clone:0610010L11 product:carbonic anhydrase 2, full insert sequence		
DE	(fragment).		
GN	CAR2.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
LN	[1]		
SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Kidney;		
RA	Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,		
RA	Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,		
RA	Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,		
RA	Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,		
RA	Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,		
RA	Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,		
RA	Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,		
RA	Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,		
RA	Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,		
RA	Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,		
RA	Muramatsu M., Hayashizaki Y.;		
RL	Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.		
[2]			
SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Kidney;		
RX	MEDLINE=22354683; PubMed=12466851;		
RA	The FANTOM Consortium,		
RA	the RIKEN Genome Exploration Research Group Phase I & II Team;		
RT	"Analysis of the mouse transcriptome based on functional annotation of		
RT	60,770 full-length cDNAs."		
RL	Nature 420:563-573 (2002).		
[3]			
SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Kidney;		
RX	MEDLINE=21085660; PubMed=11217851;		
RA	RIKEN FANTOM Consortium;		
RT	"Functional annotation of a full-length mouse cDNA collection.;"		
RL	Nature 409:685-690 (2001).		
[4]			
SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Kidney;		
RX	MEDLINE=99279253; PubMed=10349636;		
RA	Carninci P., Hayashizaki Y.;		
RT	"High-efficiency full-length cDNA cloning.;"		
RL	Meth. Enzymol. 303:19-44 (1999).		
[5]			
SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Kidney;		
RX	MEDLINE=20499374; PubMed=11042159;		
RA	Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,		
RA	Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;		
RT	"Normalization and subtraction of cap-trapper-selected cDNAs to		
RT	prepare full-length cDNA libraries for rapid discovery of new genes.;"		
RL	Genome Res. 10:1617-1630 (2000).		
[6]			
SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Kidney;		
RX	MEDLINE=20530913; PubMed=11076861;		
RA	Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,		
RA	Kono H., Akiyama J., Nishi K., Katsunai T., Tashiro H., Itoh M.,		
RA	Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,		
RA	Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,		
RA	Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,		
RA	Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,		
RA	Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;		
RT	"RIKEN integrated sequence analysis (RISA) system-384-format		
RT	sequencing pipeline with 384 multicapillary sequencer.;"		
RL	Genome Res. 10:1757-1771 (2000).		
DR	EMBL; AK002498; BAB22146.2; --		
DR	MGD; MGI:88269; Car2.		
DR	GO; GO:0004089; P:carbonate dehydratase activity; IEA.		
DR	GO; GO:0008270; P:zinc ion binding; IEA.		
DR	GO; GO:0008730; P:one-carbon compound metabolism; IEA.		
DR	InterPro; IPR001148; Euk COanhd.		
DR	Pfam; PF00194; carb anhydrase; 1.		
DR	ProDom; PD000865; Euk COanhd; 1.		
DR	PROSITE; PS00162; Euk CO2 ANHYDRASE; 1.		
FT	NON_TER 1		
SQ	SEQUENCE 192 AA; 21486 MW; E7E01D0E98692043 CRC64;		
Query Match 42.4%; Score 546.5; DB 11; Length 192;			
Best Local Similarity 59.8%; Pred. No. 7.9e-41;			
Matches 101; Conservative 28; Mismatches 39; Indels 1; Gaps 1;			
QY	70 DFDDTENKSVLRGGPVTGSYRLRQVHLHWSADHGSBHIIVDGVSYAAELHVVHNSDKY 129		
DB	1 EFDDSQDNVLRGGPLSDSYRLIQHFHFGSSDQGSSEHTVKKKYAAELHVLHWNT-KY 59		
QY	130 PSFVEAAHPDGLAVLGVFLQIGEPNSQLQKTDITDLSIKEKGKQTRFTNFDLSLLPSS 189		
DB	60 GDFGKAVQPDGLAVLGVFLKIGPASQGLQKVLALHSIKTKGKRAAFANFDPCSLPGN 119		
QY	190 WDYWTYPGSLTVPPLLESVTWIVLKPQINISSQQLAKFRSLCTAEGEA 238		
DB	120 LDYWTYPGSLTVPPLLECVTWIVREPITVSSEQMSHFRTLNFNEEGDA 168		
RESULT 13			
ID	Q8MPH8 PRELIMINARY; PRT; 243 AA.		
AC	Q8MPH8;		
DT	01-OCT-2002 (TrEMBLrel. 22, Created)		
DT	01-OCT-2002 (TrEMBLrel. 22, Last sequence update)		
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)		
DE	Carbonic anhydrase (EC 4.2.1.1).		
GN	CAL.		
OS	Riftia pachyptila (Tube worm).		
OC	Eukaryota; Metazoa; Pogonophora; Vestimentifera; Axonobranchia;		
OC	Riftiida; Riftiidae; Riftia.		
OX	NCBI_TaxID=6426;		
LN	[1]		
SEQUENCE FROM N.A.			
RA	De Cian M.C., Bailly X., Boulben S., Strub J.M., Von Dorsslaer A.,		
RA	Lallier F.H.;		
RT	"An insight into molecular and biochemical characteristics of carbonic		
RT	anhydrases from Riftia pachyptila, a symbiotic invertebrate living		
RT	under extreme conditions.;"		
RL	Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.		


```
DR EMBL; AJ439711; CAD29128.1; -.
DR GO; GO:0004089; F:carbonate dehydratase activity; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006730; P:one-carbon compound metabolism; IEA.
DR InterPro; IPR001148; Euk COanhd.
DR Pfam; PF00194; carb anhydrase; 1.
DR ProDom; PD000865; Euk COanhd; 1.
DR PROSITE; PS00162; EUK_CO2_ANHYDRASE; 1.
KW Lyase.
SQ
SEQUENCE 243 AA; 25661 MW; AAA53B407A1F4645 CRC64;

Query Match
Best Local Similarity 37.8%; Score 487.5; DB 5; Length 243;
Matches 103; Conservative 25; Mismatches 91; Indels 9; Gaps 5;

QY 5 SWGYREHNGPIHWKEFFPIADGQQQPIETKTEVKYDSSLRPLSKYDPSSAKIISNSG 64
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
3 AMDY-EANGPATWAKSFPLAAGKKQSPIDIDPASVS-KKSTSLVASYNPASNTLITNG 60
QY 65 HSFNVDFDTEKNSVLRGGLTGSYRLRQVHLHWGSAADHGHSEHIVDGVSYAAELHVVHW 124
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
61 LSFQVSVDTG-----LSGGPLGNEYKAASFHFHWSKTSAGGSEHTVAGKAYAAEAHIVHY 115
QY 125 NSDKYPSFVEAAHEPDGLAVLGVLQIGEPNSLOKITDTLDSIKEKGKQTRFT-NFDLL 183
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
116 NAAKTASFQDAVKADGLAVLATFIPQATNAGVQKLIIDLLPSVPTKGDATIPGGFDVA 175
QY 184 SLLPFSWD-YWTYPGSLTPVPLLESVTWLVKQPINISSQQLAKFRSL 230
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
176 CLLPGDSQSKYVYVPGSLTTPPCFESVTWLVKQPINISSQQLAKFRSL 223

RESULT 14
Q8K2J1
ID Q8K2J1 PRELIMINARY; PRT; 344 AA.
AC Q8K2J1
DT 01-OCT-2002 (TremBLrel. 22, Created)
DT 01-OCT-2002 (TremBLrel. 22, Last sequence update)
DE Hypothetical protein.
GN 2310047E01RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RA Strausberg R.;
RP Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC031385; AAH31385.1; -.
DR MGD; MGI:1923709; 2310047E01RIK.
DR GO; GO:0004089; F:carbonate dehydratase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006730; P:one-carbon compound metabolism; IEA.
DR InterPro; IPR001148; Euk COanhd.
DR Pfam; PF00194; carb anhydrase; 1.
DR ProDom; PD000865; Euk COanhd; 1.
DR PROSITE; PS00162; EUK_CO2_ANHYDRASE; 1.
KW Hypothetical protein.
SQ
SEQUENCE 344 AA; 38724 MW; 12CDBD6C3B45D75A CRC64;

Query Match
Best Local Similarity 33.9%; Score 436.5; DB 11; Length 344;
Matches 93; Conservative 39; Mismatches 90; Indels 9; Gaps 5;

QY 6 WGYREHNGPIHWKEFFPIADGQQQPIETKTEVKYDSSLRPLSKYDPSSAKI--ISN 62
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
32 WTYVGPAGEKWSKXYPSCGGLQSPIDILHSDILOYDASLAPLQFGYVNSVEKLNLTN 91
QY 63 SGHSFNVDFTDENKSVLRGGPLTGSYRLRQVHLHWGSAADHGHSEHIVDGVSYAAELHW 121
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
92 DGHVSRLNL----NSDMYIQGLQPHHYRAEQHLHWNENRDPHGSEHTVSGKHFAELHI 147

QY 122 VHWNSDKYPSFVEAAHEPDGLAVLGVLQIGEPNSLOKITDTLDSIKEKGKQTRFTND 181
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
148 VHYNSDLYPDFTASDKSEGLAVLAVLIEIGSANPSYDKIFSHLQHVYKVGQVLLPGFN 207
QY 182 LLSLPPS-WDYWTYPGSLTPVPLLESVTWLVKQPINISSQQLAKFRSL 231
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
208 IBE LLPSPGEYRYEGSLTTPCPTVLTMTVFRNPVQISQEQLLALETAL 258

RESULT 15
O93587
ID O93587 PRELIMINARY; PRT; 259 AA.
AC O93587
DT 01-NOV-1998 (TremBLrel. 08, Created)
DT 01-NOV-1998 (TremBLrel. 08, Last sequence update)
DE Carbonic anhydrase.
OS Platicthys flesus (European flounder).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
OC Pleuronectidae; Pleuronectidae; Platicthys.
OX NCBI_TaxID=8260;
RN [1]
RA Wright C., Cossins A.R.C.;
RP Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF093622; AAC64172.1; -.
DR HSSP; P00918; 1C1M.
DR GO; GO:0004089; F:carbonate dehydratase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006730; P:one-carbon compound metabolism; IEA.
DR InterPro; IPR001148; Euk COanhd.
DR Pfam; PF00194; carb anhydrase; 2.
DR ProDom; PD000865; Euk COanhd; 1.
SQ
SEQUENCE 259 AA; 27762 MW; DBFBF4E5E74AE6BC CRC64;

Query Match
Best Local Similarity 31.5%; Score 406; DB 13; Length 259;
Matches 95; Conservative 41; Mismatches 71; Indels 56; Gaps 8;

QY 4 LSWGYREHNGPIHWKEFFPIADGQQQPIETKTEVKYDSSLRPLSKYDPSSAKIISNS 63
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1 MSWGYAADNGPDKWADNFPVANGPRQSPIDILPGDASFDAAALKPLSLKYDPTARSILNN 60
QY 64 GHSFNVDFTDENKSVLRGGPLTGSYRLRQVHLHWGSAADHGHSEHIVDGVSYAAELHVVH 123
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
61 GHSFQVTFIDTDSSTLKDGPISGVYRLKQFHFWGACDEKGESEHTVAG----- 109
QY 124 WNSDKYPSFVEAAHEPDGLAVLGVLQIGEPNSQ-----LQKITDT-LDSIKEK-GKQTR 176
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
110 -NWCILPSSISCT-----GTNTRASEKPLANLTDSPISSEFSKLGKLTTP 153
QY 177 FTFNFDLLSLLPPSWDYWTYPGSL-TVPPLLESVTW-----IVLKQ 215
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
154 -TSRRFLTPSVPRPKASRPPLASTPPACSPGAWTTGSKKAPGLPLCWRASPLGSARE 212
QY 216 PINTISSQQLAKFRSLTCTAEGEA 238
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
213 PISVSAEQMAKFRSLLSAEGEA 235

Search completed: September 9, 2004, 15:13:55
Job time : 118 secs
```


GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 15:09:25 ; Search time 32 Seconds
(without alignments)
390.421 Million cell updates/sec

Title: US-10-069-434-1

Perfect score: 1289

Sequence: 1 MSRLSWGVRHNGPHWKEF.....QLAKERSLLCTAGEAARFL 242

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*

2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*

3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*

4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*

5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*

6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	804	62.4	261	1	US-08-015-986A-8
2	804	62.4	261	2	US-08-446-363-8
3	802	62.2	260	1	US-08-015-973-4
4	802	62.2	260	2	US-08-448-164-4
5	802	62.2	260	3	US-08-081-929-4
6	788.5	61.2	259	1	US-08-015-973-5
7	788.5	61.2	259	2	US-08-448-164-5
8	788.5	61.2	259	3	US-08-081-929-5
9	788.5	61.2	260	1	US-08-015-986A-9
10	788.5	61.2	260	2	US-08-446-363-9
11	788.5	61.2	260	3	US-08-927-128-4
12	788.5	61.2	260	4	US-09-976-594-644
13	788.5	61.2	288	2	US-08-595-868C-12
14	788.5	61.2	288	3	US-09-139-819A-12
15	788.5	61.2	288	4	US-09-750-913-12
16	748.5	58.1	259	1	US-08-015-986A-10
17	748.5	58.1	259	2	US-08-015-973-6
18	748.5	58.1	259	2	US-08-446-363-10
19	748.5	58.1	259	3	US-08-448-164-6
20	748.5	58.1	259	3	US-08-081-929-6
21	747.5	58.0	421	4	US-09-938-270B-1
22	695	53.9	201	1	US-08-469-667-18
23	695	53.9	201	4	US-09-224-110-18
24	695	53.9	201	5	PCT-US95-07289-18
25	668	51.8	261	1	US-08-015-973-9
26	668	51.8	261	2	US-08-448-164-9
27	668	51.8	261	3	US-08-081-923-9

28 452 35.1 266 1 US-08-015-986A-12 Sequence 12, Appl
29 452 35.1 266 2 US-08-446-363-12 Sequence 12, Appl
30 411.5 31.9 271 1 US-08-276-919-10 Sequence 10, Appl
31 411.5 31.9 271 1 US-08-776-088-13 Sequence 13, Appl
32 411.5 31.9 271 5 PCT-US95-09145A-13 Sequence 13, Appl
33 411.5 31.9 274 1 US-08-776-088-15 Sequence 15, Appl
34 411.5 31.9 274 5 PCT-US95-09145A-15 Sequence 15, Appl
35 411.5 31.9 325 1 US-08-276-919-4 Sequence 4, Appl
36 411.5 31.9 325 1 US-08-276-919-13 Sequence 13, Appl
37 411.5 31.9 325 1 US-08-776-088-4 Sequence 4, Appl
38 411.5 31.9 325 5 PCT-US95-09145A-18 Sequence 18, Appl
39 411.5 31.9 325 5 PCT-US95-09145A-4 Sequence 4, Appl
40 411.5 31.9 325 5 PCT-US95-09145A-18 Sequence 18, Appl
41 411.5 31.9 354 1 US-08-276-919-2 Sequence 2, Appl
42 411.5 31.9 354 1 US-08-776-088-2 Sequence 2, Appl
43 411.5 31.9 354 1 US-08-776-088-6 Sequence 6, Appl
44 411.5 31.9 354 3 US-09-325-320-2 Sequence 2, Appl
45 411.5 31.9 354 3 US-09-585-109-2 Sequence 2, Appl

ALIGNMENTS

RESULT 1

US-08-015-986A-8

; Sequence 8, Application US/08015986A

; Patent No. 5532123

; GENERAL INFORMATION:

; APPLICANT: Schlessinger, Joseph

; TITLE OF INVENTION: NOVEL RECEPTOR-TYPE PHOSPHOTYROSINE

; TITLE OF INVENTION: PHOSPHATASE-GAMMA

; NUMBER OF SEQUENCES: 14

; CORRESPONDENCE ADDRESS:

; ADDRESS: PENNIE & EDMONDS

; STREET: 1155 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10036

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/015.986A

; FILING DATE: 10-FEB-1993

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Mistrock, S. Leslie

; REGISTRATION NUMBER: 18,872

; REFERENCE/DOCKET NUMBER: 7683-028

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-790-9090

; TELEFAX: 212-869-8864/9741

; TELEX: 66141 PENNIE

; INFORMATION FOR SEQ ID NO: 8:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 261 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: unknown

; MOLECULE TYPE: protein

US-08-015-986A-8

Query Match 62.4%; Score 804; DB 1; Length 261;

Best Local Similarity 59.4%; Pred. No. 1.9e-76;

Matches 142; Conservative 44; Mismatches 53; Indels 0; Gaps 0;

Qy 1 MSRLSWGVRHNGPHWKEFFPIADGQDQSEIETKVKYDSSLRPLSIKYDPSAKII 60

Db 1 MASPDWGVDDKNGPQWQSKLPIANGNNQSPVDIKTKTKDTSKLPISVSNPATAKEI 60


```

; GENERAL INFORMATION:
; APPLICANT: Schllessinger, Joseph
; TITLE OF INVENTION: A NOVEL RECEPTOR-TYPE PHOSPHOTYROSINE
; TITLE OF INVENTION: PHOSPHATASE-BETA
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/448,164
; FILING DATE: 24-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/015,973
; FILING DATE: 10-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,972
; REFERENCE/DOCKET NUMBER: 7683-021
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)790-9090
; TELEFAX: (212)869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 260 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-448-164-4

Query Match 62.2%; Score 802; DB 2; Length 260;
Best Local Similarity 60.3%; Pred. No. 3.1e-76;
Matches 141; Conservative 43; Mismatches 50; Indels 0; Gaps 0;

QY 6 WGYREHNGPIHWKEFFPIADGQSPPIETKTEVKYDSSLRPLSIKYPSPSSAKIISNSGH 65
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 5 WGYDDKNGPEQWSKLYPIANGNNQSPVDIKTSETKHTSLKPISVSNPATAKEIINVGH 64
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 66 SFNVDFDITENKSVLRGGPLTGSYRLRQVHLHWGSDHGHSEHIVDGVSYAAELHVVHWN 125
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 65 SFHNVFENDNRNLSVLKGGPFSDSYRLQFPHFWGSTNEHSGSEHTVDGVKYSALHVAHWN 124
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 126 SDKYPSFVEAAHEPDGLAVLGVFLQIGPNLSQLOKITDTLDSIKKKGKQTRFTNFDLLSL 185
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 125 SAKYSSLAERASKADGLAVIGVMKVGEANPKLQKVLDAIQAITYKGRAPTNFDPSTL 184
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 186 LPPSWDYWTYPGSLTPPPLLESVTWIVLKQPINISSQQLAKFRSLCTAEGEAA 239
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 185 LPSSLDFTYTPGSLTHPLPYESTVTWICKESISVSSEQLAQFRSLLSNVGDN 238
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 5
US-08-081-929-4
; Sequence 4, Application US/08081929
; Patent No. 6160090
; GENERAL INFORMATION:
; APPLICANT: Schllessinger, Joseph
; APPLICANT: Barnea, Gilad
; APPLICANT: Grumet, Martin H.
; APPLICANT: Margolis, Richard U.
; TITLE OF INVENTION: A NEW CLASS OF RTPases: THEIR
; TITLE OF INVENTION: STRUCTURAL DOMAINS AND LIGANDS

; GENERAL INFORMATION:
; APPLICANT: Schllessinger, Joseph
; TITLE OF INVENTION: A NOVEL RECEPTOR-TYPE PHOSPHOTYROSINE
; TITLE OF INVENTION: PHOSPHATASE-BETA
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

; GENERAL INFORMATION:
; APPLICANT: Schllessinger, Joseph
; TITLE OF INVENTION: A NOVEL RECEPTOR-TYPE PHOSPHOTYROSINE
; TITLE OF INVENTION: PHOSPHATASE-BETA
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/081,929
; FILING DATE: 23-JUN-1993
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30742
; REFERENCE/DOCKET NUMBER: 7683-041-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 260 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-081-929-4

Query Match 62.2%; Score 802; DB 3; Length 260;
Best Local Similarity 60.3%; Pred. No. 3.1e-76;
Matches 141; Conservative 43; Mismatches 50; Indels 0; Gaps 0;

QY 6 WGYREHNGPIHWKEFFPIADGQSPPIETKTEVKYDSSLRPLSIKYPSPSSAKIISNSGH 65
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 5 WGYDDKNGPEQWSKLYPIANGNNQSPVDIKTSETKHTSLKPISVSNPATAKEIINVGH 64
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 66 SFNVDFDITENKSVLRGGPLTGSYRLRQVHLHWGSDHGHSEHIVDGVSYAAELHVVHWN 125
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 65 SFHNVFENDNRNLSVLKGGPFSDSYRLQFPHFWGSTNEHSGSEHTVDGVKYSALHVAHWN 124
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 126 SDKYPSFVEAAHEPDGLAVLGVFLQIGPNLSQLOKITDTLDSIKKKGKQTRFTNFDLLSL 185
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 125 SAKYSSLAERASKADGLAVIGVMKVGEANPKLQKVLDAIQAITYKGRAPTNFDPSTL 184
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 186 LPPSWDYWTYPGSLTPPPLLESVTWIVLKQPINISSQQLAKFRSLCTAEGEAA 239
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 185 LPSSLDFTYTPGSLTHPLPYESTVTWICKESISVSSEQLAQFRSLLSNVGDN 238
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 6
US-08-015-973-5
; Sequence 5, Application US/08015973
; Patent No. 5604094
; GENERAL INFORMATION:
; APPLICANT: Schllessinger, Joseph
; TITLE OF INVENTION: A NOVEL RECEPTOR-TYPE PHOSPHOTYROSINE
; TITLE OF INVENTION: PHOSPHATASE-BETA
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

ug-10-069-434-1.rai

ug-10-069-434-1.rai

```

Query Match          61.2%; Score 788.5; DB 3; Length 259;
Best Local Similarity 61.6%; Pred. No. 8e-75;
Matches 143; Conservative 33; Mismatches 55; Indels 1; Gaps 1;

QY      6 WGYREHNGPIHWKEFFFIADGQQSPFIEIKTEVKVYDSSLRPLSIKYDPSSAKIISNSGH 65
       ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      5 WGYRGKHGPHEWHKDFPIAKGERQSPVDIDTHTAKYPDSLKLPSLSVSVDQAATSLRLNNGH 64
       :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY     66 SFNVDFDDTENKSVLRCGGPGLTGSRYLROVHLHWGSADDHGSEHVLDGVSYAAELHVVHNW 125
       :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db     65 AFNFVEDDSQDAVKLVKGPGDLGTIRLIQHFWHWSLDCQGSGEHTVDDKKYAEEHLVHNW 124
       :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY    126 SDKYPSFVEAHPDCLAVLGVLQIGEPNSQLKITDTLDSIKEKGKOTRFTNFDLLSL 185
       ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db   125 T-KYGDFGKAVQPDLAVLGIFLKVGSAKPGLQKVWDVLDISKTKGSAADFTNFPDRL 183
       :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY    186 LPESDWITYPGSLTVPPLESVTWIVLKOPINISSQOLAKFRSLLCTAEGE 237
       ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db   184 LPESLDWTYPGSLTPPLELCVTWIVLKEPISVSSEQVLKFKLNFNENGE 235

RESULT 10
US-08-446-363-9
; Sequence 9, Application US/08446363
; Patent No. 5891700
; GENERAL INFORMATION:
; APPLICANT: Schllessinger, Joseph
; TITLE OF INVENTION: NOVEL RECEPTOR-TYPE PHOSPHOTYROSINE
; TITLE OF INVENTION: PHOSPHATASE-GAMMA
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,363
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/015,986
; FILING DATE: 10-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7683-028
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEX: 212-869-8864/9741
; INFORMATION FOR SEQ ID NO: 9:
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-446-363-9

Query Match          61.2%; Score 788.5; DB 2; Length 260;
Best Local Similarity 61.6%; Pred. No. 8e-75;
Matches 143; Conservative 33; Mismatches 55; Indels 1; Gaps 1;

QY      6 WGYREHNGPIHWKEFFFIADGQQSPFIEIKTEVKVYDSSLRPLSIKYDPSSAKIISNSGH 65
       ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      5 WGYRGKHGPHEWHKDFPIAKGERQSPVDIDTHTAKYPDSLKLPSLSVSVDQAATSLRLNNGH 64
       :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY     66 SFNVDFDDTENKSVLRCGGPGLTGSRYLROVHLHWGSADDHGSEHVLDGVSYAAELHVVHNW 125
       :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db     65 AFNFVEDDSQDAVKLVKGPGDLGTIRLIQHFWHWSLDCQGSGEHTVDDKKYAEEHLVHNW 124
       :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY    126 SDKYPSFVEAHPDCLAVLGVLQIGEPNSQLKITDTLDSIKEKGKOTRFTNFDLLSL 185
       ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db   125 T-KYGDFGKAVQPDLAVLGIFLKVGSAKPGLQKVWDVLDISKTKGSAADFTNFPDRL 183
       :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY    186 LPESDWITYPGSLTVPPLESVTWIVLKOPINISSQOLAKFRSLLCTAEGE 237
       ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db   184 LPESLDWTYPGSLTPPLELCVTWIVLKEPISVSSEQVLKFKLNFNENGE 235

RESULT 9
US-08-015-986A-9
; Sequence 9, Application US/08015986A
; Patent No. 5532123
; GENERAL INFORMATION:
; APPLICANT: Schllessinger, Joseph
; TITLE OF INVENTION: NOVEL RECEPTOR-TYPE PHOSPHOTYROSINE
; TITLE OF INVENTION: PHOSPHATASE-GAMMA
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/015,986A
; FILING DATE: 10-FEB-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7683-028
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEX: 212-869-8864/9741
; INFORMATION FOR SEQ ID NO: 9:
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-015-986A-9

```


Qy	66	SFNVDFDDTENKSVLRRGGPLTGSYRLRVQLHWSGADHGSEHIVDGVSYAAELHVVHN	125
Dd	65	AFFVFDDSQDKAVLKGGPLDGTLYLIQFHFWHSGLDQGSEHTVDKKKYAAELHLVHN	124
Qy	126	SDKPSFVEAAHEHPDGLAVLGFVLIGBPNSQLQKITDTLDLSIEKKGQTFNFLLSL	185
Dd	125	T-KYGFDFGKAVQQPDGLAVLGIFLKVSGAKPGLQKVVVDLDSIKTKGSADTFNFPRLG	183
Qy	186	LPPSWDYWTYPGSLTVPPPLLESVTWLVLKQPINSSOQLAKFRSILLCTAEGE	237
Dd	184	LPBSLDYWTYPGSLTPPLLECMTWLVLEKEPTISVSSEQVLFRKLNFNGGE	235

RESULT 11
 US-08-927-128-4
 : Sequence 4, Application US/08927128
 : Patent No. 6127150
 : GENERAL INFORMATION:
 : APPLICANT: Coolidge, Thomas
 : APPLICANT: Wagner, Fred
 : APPLICANT: ven Heeke, Gino
 : APPLICANT: Schuster, Sheldon
 : APPLICANT: Stout, Jay
 : APPLICANT: Wylie, Dwane
 : TITLE OF INVENTION: PURIFICATION DIRECTED CLOSING OF PEPTIDES
 : NUMBER OF SEQUENCES: 28
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Merchant & Gould
 : STREET: 3100 No. 6127150west Center, 90 S. 7th Street
 : CITY: Minneapolis
 : STATE: MN
 : COUNTRY: U.S.A.
 : ZIP: 55402
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Diskette
 : COMPUTER: IBM Compatible
 : OPERATING SYSTEM: DOS
 : SOFTWARE: FastSEQ Version 1.5
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/927,128
 : FILING DATE: 05-SEP-1997
 : CLASSIFICATION: 435
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: 08/680,004
 : FILING DATE: 15-JUL-1995
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Carter, Charles G
 : REGISTRATION NUMBER: 35,093
 : REFERENCE/DOCKET NUMBER: 8648.2USD1
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: 612/332-5300
 : TELEFAX: 612/332-9081
 : TELEX:
 : INFORMATION FOR SEQ ID NO: 4:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 260 amino acids
 : TYPE: amino acid
 : STRANDEDNESS: single
 : TOPOLOGY: linear
 : MOLECULE TYPE: protein
 : HYPOTHETICAL: NO
 : ANTI-SENSE: NO
 : FRAGMENT TYPE: internal
 : ORIGINAL SOURCE:
 : US-08-927-128-4

```

Db      5  WGYGKHNGPEHHKDPPIAKGERQSPVDIDTHTAKYDPSLKPLSVSDQATSLRLNNGH  64
Qy      66  SENVDPDPTENKSVLRGGPLTGSYRLRQVHLHWGSAADHGSEHIVDGSYAAELHVVVHN  125
Db      65  AFNVFEDDSQDKAVLKGPPDGGYIRLLQFHHWGLDGCQSEHTVDKKYAAELHLVHN  124
Qy      126  SDKYPSVEAAHPDGLAVLGVFIQIGEPNSQLQKITDILDSIKEKGKQTRFTNFDLLSL  185
Db      125  T-KYGDCKAVQPDGLAVLGIFLKVGSAKPGLQKVVDVLDLSIKTKGSADFTNFDPRGL  183
Qy      186  LPSWDYWTYPGSLTVPLLESVTWIVLKOPINISSQOLAKFRSLLLCTARGE  237
Db      184  LPESLDYWTYPGSLTTPPLECVTWIVLKPISVSSEQVLKFRKLNFNGE  235

RESULT 12
US-09-976-594-644
; Sequence 644. Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: P8-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 644
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 2742913CD1
US-09-976-594-644

```

```

Query Match      61.2%; Score 788.5; DB 4; Length 260;
Best Local Similarity 61.6%; Pred. No. 8e-75; 55; Indels 1; Gaps 1;
Matches 143; Conservative 33; Mismatches 33;

6 WGYREHNGPIHWKEPFPIADGQQSPPIEIKTKEVKYDSSLRPLSIKYDPSSAKIIISNGH 65
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5 WGYGKHNGPEHWHKDFPIAKGERQSPVIDHTHAKYDPSLKPFLSVSYDQATSLRILNNGH 64
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
66 SFNVDFDPTENKNSVLVRGGLTGSYKRLRVHLHWGSADHGHSEHIVDGSYAAELHVVHWN 125
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
65 AFNVFEDDSQDKAVLKGGLPDLGTYRLIOPHFHWGSLDGGSEHTVDKKKYAAELHLVHWN 124
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
126 SDKYSFVEAAHEPGLAVLGVFLGIPNSOLOKITDLDLSIKEKGKQTRTNFDLLSL 185
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
125 T-KYGDGKAVQOPGLAVLGLIFLKVGSAAKQGLQKVQDVLDLSIKYKGSADFTNDFPRGL 183
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
186 LPSPDYWTYPGSLTVPPLLESVTVILVKQPINISSQQLAKFRSLTCTAE 237
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
184 LPESLDYWTYPGSLTTPPLECTVTVILKEPISVSSEQVLKFKLNFNGE 235
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
US-08-595-868C-12
; Sequence 12, Application US/08595868C
; Patent No. 5962270
; GENERAL INFORMATION:
; APPLICANT: Wagner, Fred
; APPLICANT: Stout, Jay
; APPLICANT: Henriksen, Dennis
; APPLICANT: Partidge, Bruce
; APPLICANT: Holmquist, Bart
; APPLICANT: Frank, Julie
; TITLE OF INVENTION: RECOMBINANT PREPARATION OF CALCITONIN FRAGMENTS AND US
; NUMBER OF SEQUENCES: 51

```

```

RESULT 13
US-08-595-868C-12
; Sequence 12, Application US/08595868C
; Patent No. 5962270
; GENERAL INFORMATION:
; APPLICANT: Wagner, Fred
; APPLICANT: Stout, Jay
; APPLICANT: Henriksen, Dennis
; APPLICANT: Partridge, Bruce
; APPLICANT: Holmquist, Bart
; APPLICANT: Frank, Julie
; TITLE OF INVENTION: RECOMBINANT PREPARATION OF CALCITONIN FRAGMENTS AND US
; NUMBER OF SEQUENCES: 51

```

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant & Gould
; STREET: 3100 No. 5962270west Center, 90 S. 7th Street
; CITY: Minneapolis
; STATE: MN
; COUNTRY: U.S.A.
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: Fast-SEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/595,868C
; FILING DATE: 06-FEB-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Carter, Charles G
; REGISTRATION NUMBER: 35,093
; REFERENCE/DOCKET NUMBER: 8648.59US01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612/332-5300
; TELEFAX: 612/332-9081
; TELEX:
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 288 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: C-terminal
; ORIGINAL SOURCE:
; US-08-595-868C-12

Query Match 61.2%; Score 788.5; DB 2; Length 288;
Best Local Similarity 61.6%; Pred. No. 9.4e-75;
Matches 143; Conservative 33; Mismatches 55; Indels 1; Gaps 1;
QY 6 WGYRHNHGPPIHKEFPPIADGQQSPPIETKTEVKYDSSLRPLSIKYDPSSAKIISNSGH 65
Db 5 WGYKHNGPEHWHKDFPIAKGERQSPVDIDHTAKYDPSLKLPSVSYDQATSLRIINNGH 64
QY 66 SFNVDFDPTENKSVLRGGPLTGSYRLRQVHLHWSADHDGSEHIVDGVSYAAELHVVHWN 125
Db 65 AFNVEFDDSDQKAVLKGGLDGTLYLIQFPHWGLDGGSEHTVDDKKYAAELHVVHWN 124
QY 126 SDKYPSFVEAAHEPDLAVLGVLQIGEPNSQLKIDTLDLSIKKKGKQTRFTNFDLLSL 185
Db 125 T-KYDGFKAVQPPDGLAVLGIFLKVGSAPKGLQKVDVLDLSIKTKGKSADFTNFDPRGL 183
QY 186 LPPSDYWTYPGSLTVPPLLESVTWVLKOPINISSQQLAKFRSLCTAEGE 237
Db 184 LPESLDYWTYPGSLTTPPLLECVTWVLKEPISVSSEQVLFKRLNFNGEGE 235

RESULT 14
US-09-139-819A-12
; Sequence 12, Application US/09139819A
; Patent No. 6251635
; GENERAL INFORMATION:
; APPLICANT: WAGNER, Fred W.
; APPLICANT: STOUT, Jay S.
; APPLICANT: HENRIKSEN, Dennis B.
; APPLICANT: PARTRIDGE, Bruce E.
; APPLICANT: HOLMQUIST, Bart
; APPLICANT: FRANK, Julie A.
; TITLE OF INVENTION: RECOMBINANT PREPARATION OF CALCITONIN

; TITLE OF INVENTION: FRAGMENTS AND USE THEREOF IN THE PREPARATION OF CALCITONIN
; TITLE OF INVENTION: AND RELATED ANALOGS
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/139,819A
; FILING DATE: 25-AUG-1998
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/595,868
; FILING DATE: 06-FEB-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 089187/0144
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 288 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-139-819A-12

Query Match 61.2%; Score 788.5; DB 3; Length 288;
Best Local Similarity 61.6%; Pred. No. 9.4e-75;
Matches 143; Conservative 33; Mismatches 55; Indels 1; Gaps 1;
QY 6 WGYRHNHGPPIHKEFPPIADGQQSPPIETKTEVKYDSSLRPLSIKYDPSSAKIISNSGH 65
Db 5 WGYKHNGPEHWHKDFPIAKGERQSPVDIDHTAKYDPSLKLPSVSYDQATSLRIINNGH 64
QY 66 SFNVDFDPTENKSVLRGGPLTGSYRLRQVHLHWSADHDGSEHIVDGVSYAAELHVVHWN 125
Db 65 AFNVEFDDSDQKAVLKGGLDGTLYLIQFPHWGLDGGSEHTVDDKKYAAELHVVHWN 124
QY 126 SDKYPSFVEAAHEPDLAVLGVLQIGEPNSQLKIDTLDLSIKKKGKQTRFTNFDLLSL 185
Db 125 T-KYDGFKAVQPPDGLAVLGIFLKVGSAPKGLQKVDVLDLSIKTKGKSADFTNFDPRGL 183
QY 186 LPPSDYWTYPGSLTVPPLLESVTWVLKOPINISSQQLAKFRSLCTAEGE 237
Db 184 LPESLDYWTYPGSLTTPPLLECVTWVLKEPISVSSEQVLFKRLNFNGEGE 235

RESULT 15
US-09-750-913-12
; Sequence 12, Application US/09750913
; Patent No. 6410707
; GENERAL INFORMATION:
; APPLICANT: WAGNER, Fred W.
; APPLICANT: STOUT, Jay S.
; APPLICANT: HENRIKSEN, Dennis B.
; APPLICANT: PARTRIDGE, Bruce E.
; APPLICANT: HOLMQUIST, Bart
; APPLICANT: FRANK, Julie A.
; TITLE OF INVENTION: RECOMBINANT PREPARATION OF CALCITONIN
; FRAGMENTS AND USE THEREOF IN THE PREPARATION OF CALCITONIN
; AND RELATED ANALOGS
; NUMBER OF SEQUENCES: 51

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 15:12:00 ; Search time 135 Seconds
(without alignments)
574.866 Million cell updates/sec

Title: US-10-069-434-1

Perfect score: 1289

Sequence: 1 MSRLSWGVRHNGPIHWKEF.....QLAKERSLLCTAGGAAAF 242

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1335176 seqs, 320689617 residues

Total number of hits satisfying chosen parameters: 1335176

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pap.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pap.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pap.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pap.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pap.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pap.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pap.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pap.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pap.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pap.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pap.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pap.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pap.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pap.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pap.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pap.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pap.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1289	100.0	242	14	US-10-069-434-1
2	1289	100.0	262	12	US-10-433-802-1
3	1289	100.0	262	15	US-10-108-260A-3082
4	1289	100.0	274	12	US-10-220-120-215
5	807	62.6	255	15	US-10-099-322-304
6	807	62.6	255	15	US-10-044-564-304
7	804	62.4	261	9	US-09-802-674-4
8	804	62.4	261	9	US-09-981-353-80
9	804	62.4	261	14	US-10-235-934-24
10	804	62.4	261	16	US-10-408-765A-389
11	804	62.4	263	14	US-10-106-698-4637
12	802	62.2	260	10	US-09-983-000A-26
13	802	62.2	260	13	US-10-000-954-4
14	788.5	61.2	259	13	US-10-000-954-5
15	788.5	61.2	260	9	US-09-981-353-117

16 788.5 61.2 294 9 US-09-925-299-945
17 788.5 61.2 294 10 US-09-925-299-945
18 748.5 58.1 259 10 US-09-983-000A-25
19 748.5 58.1 259 13 US-10-000-954-6
20 748 58.0 242 15 US-10-231-913-284
21 747.5 58.0 260 12 US-10-363-616-250
22 747.5 58.0 421 9 US-09-938-270B-1
23 695 53.9 201 12 US-09-988-292-18
24 695 53.9 201 16 US-10-776-601-18
25 668 51.8 261 13 US-10-000-954-9
26 506 39.3 170 16 US-10-038-854-106
27 453 35.1 291 15 US-10-231-913-146
28 453 35.1 291 15 US-10-231-913-147
29 451 35.0 289 14 US-10-074-475-239
30 451 35.0 289 15 US-10-231-913-145
31 451 35.0 290 15 US-10-231-913-143
32 451 35.0 290 15 US-10-231-913-144
33 411.5 31.9 354 15 US-10-295-027-296
34 411.5 31.9 354 16 US-10-648-593-216
35 411 31.9 81 16 US-10-363-829-298
36 376.5 29.2 337 10 US-09-946-374-423
37 376.5 29.2 337 10 US-09-983-000A-27
38 376.5 29.2 337 12 US-10-206-915-268
39 376.5 29.2 337 12 US-10-199-670-268
40 376.5 29.2 337 12 US-10-201-858-268
41 376.5 29.2 337 12 US-10-205-890-268
42 376.5 29.2 337 12 US-10-208-024-268
43 376.5 29.2 337 12 US-10-201-853-268
44 376.5 29.2 337 12 US-10-063-745-74
45 376.5 29.2 337 12 US-10-063-512-74

ALIGNMENTS

RESULT 1

US-10-069-434-1
; Sequence 1, Application US/10069434
; Publication No. US20030121061A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: THORNTON, Michael
; APPLICANT: RAMKUMAR, Jayalaxmi
; APPLICANT: TRIBOULEY, Catherine M.
; APPLICANT: YUE, Henry
; APPLICANT: NGUYEN, Dannel B.
; APPLICANT: YAO, Monique G.
; APPLICANT: PATTERSON, Chandra
; APPLICANT: GANDHI, Ameena R.
; APPLICANT: BURFORD, Neil
; APPLICANT: THANGAVELU, Kavitha
; APPLICANT: BAUGHN Mariah R.
; TITLE OF INVENTION: HUMAN LYASES
; FILE REFERENCE: PI-0137 PCT
; CURRENT APPLICATION NUMBER: US/10/069,434
; CURRENT FILING DATE: 2002-02-20
; PRIOR APPLICATION NUMBER: 60/213,383; 60/215,544; 60/222,818
; PRIOR FILING DATE: 2000-06-23; 2000-06-30; 2000-08-04
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PERL Program
; SEQ ID NO 1
; LENGTH: 242
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030121061A1 6338333CD1
US-10-069-434-1

Query Match 100.0%; Score 1289; DB 14; Length 242;
Best Local Similarity 100.0%; Pred. No. 3.9e-125;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MSRLSWGYNHNGPIHWKEFFPIADGDOOSPIEIKTKVKYDSSLRPLSIKYDPSSAKII 60
|||
DB 1 MSRLSWGYNHNGPIHWKEFFPIADGDOOSPIEIKTKVKYDSSLRPLSIKYDPSSAKII 60
|||
QY 61 SNSGHSFNVD FDDTENKSVLRGGPLTGSYRLRQVHLHWGSADDDHSGSEHIVDGVSYAAELH 120
|||
DB 61 SNSGHSFNVD FDDTENKSVLRGGPLTGSYRLRQVHLHWGSADDDHSGSEHIVDGVSYAAELH 120
|||
QY 121 VVHNSDKYPSFVEAAHEPDGLAVLGVLQIGEPNSQLQKITDTLDSIKEKGKQTRFTNF 180
|||
DB 121 VVHNSDKYPSFVEAAHEPDGLAVLGVLQIGEPNSQLQKITDTLDSIKEKGKQTRFTNF 180
|||
QY 181 DLLSLPPSWDYWTYPGSLTVPPLESVTWIVLKQPINISSQOLAKFRSLLCTAAGEAAA 240
|||
DB 181 DLLSLPPSWDYWTYPGSLTVPPLESVTWIVLKQPINISSQOLAKFRSLLCTAAGEAAA 240
|||
QY 241 FL 242
||
DB 241 FL 242
```

RESULT 2

```
US-10-433-802-1
; Sequence 1, Application US/10433802
; Publication No. US20040063115A1
; GENERAL INFORMATION:
; APPLICANT: TANG, Y. Tom; GRIFFIN, Jennifer A.;
; APPLICANT: YUE, Henry; LEE, Ernestine A.;
; APPLICANT: BAUGHN, Mariah R.; DUGGAN, Brendan M.;
; APPLICANT: CHAMLA, Narinder K.; LEE, Sally;
; APPLICANT: RAMKUMAR, Jayalaxmi; WARREN, Bridget A.;
; APPLICANT: GANDHI, Ameena R.; LU, Dying Aina M.;
; APPLICANT: LU, Yan; YAO, Monique G.;
; APPLICANT: DING, Li; TRIBOULEY, Catherine M.;
; APPLICANT: SANJANWALA, Madhusudan M.;
; APPLICANT: ARVIZU, Chandra S.; JACKSON, Jennifer L.
```

```
; TITLE OF INVENTION: ENZYMS
```

```
; FILE REFERENCE: PI-0316 PCT
```

```
; CURRENT APPLICATION NUMBER: US/10/433,802
```

```
; CURRENT FILING DATE: 2003-06-04
```

```
; PRIOR APPLICATION NUMBER: PCT/US01/47432
```

```
; PRIOR FILING DATE: 2001-12-04
```

```
; PRIOR APPLICATION NUMBER: US 60/251,824
```

```
; PRIOR FILING DATE: 2000-12-07
```

```
; PRIOR APPLICATION NUMBER: US 60/254,312
```

```
; PRIOR FILING DATE: 2000-12-08
```

```
; PRIOR APPLICATION NUMBER: US 60/255,773
```

```
; PRIOR FILING DATE: 2000-12-14
```

```
; PRIOR APPLICATION NUMBER: US 60/256,188
```

```
; PRIOR FILING DATE: 2000-12-15
```

```
; PRIOR APPLICATION NUMBER: US 60/255,940
```

```
; PRIOR FILING DATE: 2000-12-15
```

```
; PRIOR APPLICATION NUMBER: US 60/257,488
```

```
; PRIOR FILING DATE: 2000-12-21
```

```
; PRIOR APPLICATION NUMBER: US 60/262,839
```

```
; PRIOR FILING DATE: 2001-01-19
```

```
; PRIOR APPLICATION NUMBER: US 60/264,402
```

```
; PRIOR FILING DATE: 2001-01-26
```

```
; NUMBER OF SEQ ID NOS: 36
```

```
; SOFTWARE: PERL Program
```

```
; SEQ ID NO 1
```

```
; LENGTH: 262
```

```
; TYPE: PRT
```

```
; ORGANISM: Homo sapiens
```

```
; NAME/KEY: misc feature
```

```
; OTHER INFORMATION: Incyte ID No. US20040063115A1 8159895CD1
```

```
US-10-433-802-1
```

```
Query Match
```

```
Best Local Similarity 100.0%; Score 1289; DB 12; Length 262;
```

```
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 MSRLSWGYNHNGPIHWKEFFPIADGDOOSPIEIKTKVKYDSSLRPLSIKYDPSSAKII 60
|||
DB 1 MSRLSWGYNHNGPIHWKEFFPIADGDOOSPIEIKTKVKYDSSLRPLSIKYDPSSAKII 60
|||
QY 61 SNSGHSFNVD FDDTENKSVLRGGPLTGSYRLRQVHLHWGSADDDHSGSEHIVDGVSYAAELH 120
|||
DB 61 SNSGHSFNVD FDDTENKSVLRGGPLTGSYRLRQVHLHWGSADDDHSGSEHIVDGVSYAAELH 120
|||
QY 121 VVHNSDKYPSFVEAAHEPDGLAVLGVLQIGEPNSQLQKITDTLDSIKEKGKQTRFTNF 180
|||
DB 121 VVHNSDKYPSFVEAAHEPDGLAVLGVLQIGEPNSQLQKITDTLDSIKEKGKQTRFTNF 180
|||
QY 181 DLLSLPPSWDYWTYPGSLTVPPLESVTWIVLKQPINISSQOLAKFRSLLCTAAGEAAA 240
|||
DB 181 DLLSLPPSWDYWTYPGSLTVPPLESVTWIVLKQPINISSQOLAKFRSLLCTAAGEAAA 240
|||
QY 241 FL 242
||
DB 241 FL 242
```

RESULT 3

```
US-10-108-260A-3082
; Sequence 3082, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
```

```
; TITLE OF INVENTION: No. US20040005560A1el full length cdna
```

```
; FILE REFERENCE: H1-A0106
```

```
; CURRENT APPLICATION NUMBER: US/10/108,260A
```

```
; CURRENT FILING DATE: 2002-03-27
```

```
; NUMBER OF SEQ ID NOS: 5458
```

```
; SOFTWARE: Patent In Ver. 2.1
```

```
; SEQ ID NO 3082
```

```
; LENGTH: 262
```

```
; TYPE: PRT
```

```
; ORGANISM: Homo sapiens
```

```
US-10-108-260A-3082
```

```
Query Match
```

```
Best Local Similarity 100.0%; Score 1289; DB 15; Length 262;
```

```
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 MSRLSWGYNHNGPIHWKEFFPIADGDOOSPIEIKTKVKYDSSLRPLSIKYDPSSAKII 60
```

```
DB 1 MSRLSWGYNHNGPIHWKEFFPIADGDOOSPIEIKTKVKYDSSLRPLSIKYDPSSAKII 60
```

```
QY 61 SNSGHSFNVD FDDTENKSVLRGGPLTGSYRLRQVHLHWGSADDDHSGSEHIVDGVSYAAELH 120
```

```
DB 61 SNSGHSFNVD FDDTENKSVLRGGPLTGSYRLRQVHLHWGSADDDHSGSEHIVDGVSYAAELH 120
```

```
QY 121 VVHNSDKYPSFVEAAHEPDGLAVLGVLQIGEPNSQLQKITDTLDSIKEKGKQTRFTNF 180
```

```
DB 121 VVHNSDKYPSFVEAAHEPDGLAVLGVLQIGEPNSQLQKITDTLDSIKEKGKQTRFTNF 180
```

```
QY 181 DLLSLPPSWDYWTYPGSLTVPPLESVTWIVLKQPINISSQOLAKFRSLLCTAAGEAAA 240
```

```
DB 181 DLLSLPPSWDYWTYPGSLTVPPLESVTWIVLKQPINISSQOLAKFRSLLCTAAGEAAA 240
```

```
QY 241 FL 242
```

```
DB 241 FL 242
```

RESULT 4

```
US-10-220-120-215
```

```
; Sequence 215, Application US/10220120
```

```
; Publication No. US20040048253A1
```

```
; GENERAL INFORMATION:
```

```
; APPLICANT: INCYTE GENOMICS, INC.
```

```
; APPLICANT: FANZER, Scott R.
```

```
; APPLICANT: SPIRO, Peter A.
```

```
; APPLICANT: BANVILLE, Steven C.
```

```
; APPLICANT: SHAH, Purvi
; APPLICANT: CHALUP, Michael S.
; APPLICANT: CHANG, Simon C.
; APPLICANT: CHEN, Alice
; APPLICANT: D'SA, Steven A.
; APPLICANT: AMSHEY, Stefan
; APPLICANT: DAHL, Christopher R.
; APPLICANT: DAM, Tam C.
; APPLICANT: DANIELS, Susan E.
; APPLICANT: DUFOUR, Gerard E.
; APPLICANT: FLORES, Vincent
; APPLICANT: FONG, Willy T.
; APPLICANT: GREENAWALT, Lila B.
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: JONES, Anissa L.
; APPLICANT: LIU, Tommy F.
; APPLICANT: ROSEBERRY, Ann M.
; APPLICANT: ROSEN, Bruce H.
; APPLICANT: RUSSO, Frank D.
; APPLICANT: STOCKREHER, Theresa K.
; APPLICANT: DAFFO, Abel
; APPLICANT: WRIGHT, Rachel J.
; APPLICANT: YAP, Pierre E.
; APPLICANT: YU, Jimmy Y.
; APPLICANT: BRADLEY, Diana L.
; APPLICANT: BRATCHER, Shawn R.
; APPLICANT: CHEN, Wensheng
; APPLICANT: COHEN, Howard J.
; APPLICANT: HODGSON, David M.
; APPLICANT: LINCOLN, Stephen E.
; APPLICANT: JACKSON, Stuart
; TITLE OF INVENTION: MOLECULES FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PT-1113 PCT
; CURRENT APPLICATION NUMBER: US/10/220,120
; PRIOR FILING DATE: 2002-08-26
; PRIOR APPLICATION NUMBER: 60/184,777; 60/184,797; 60/184,698; 60/184,770; 60/184,774;
; 60/184,693; 60/184,771; 60/184,813; 60/184,773; 60/184,776;
; 60/184,769; 60/184,768; 60/184,837; 60/184,697; 60/184,841;
; 60/184,772; 60/185,213; 60/185,216; 60/204,863; 60/205,221;
; 60/204,815; 60/203,785; 60/204,821; 60/204,908; 60/204,226;
; 60/204,525; 60/205,285; 60/205,232; 60/205,323; 60/205,287;
; 60/205,324; 60/205,286
; PRIOR FILING DATE: 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24;
; 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24;
; 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24;
; 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24;
; 2000-05-17; 2000-05-12; 2000-05-16; 2000-05-16; 2000-05-15;
; 2000-05-16; 2000-05-17; 2000-05-16; 2000-05-17; 2000-05-17;
; 2000-05-17; 2000-05-17
; NUMBER OF SEQ ID NOS: 422
; SOFTWARE: PERL Program
; SEQ ID NO 215
; LENGTH: 274
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte to No. US20040048253A1 LI:090574.1.orf3:2000FBB01
US-10-220-120-215

Query Match 100.0%; Score 1289; DB 12; Length 274;
Best Local Similarity 100.0%; Pred. No. 4.6e-125;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSRLSGYREHNGPIHWKEFFFIADGQSPTEIKTEKVYDSSLRPLSIKYDPSSAKII 60
Db 13 MSRLSGYREHNGPIHWKEFFFIADGQSPTEIKTEKVYDSSLRPLSIKYDPSSAKII 72
QY 61 SNSGHSFNVDFTDENKSVLRGGLTGSLRQVHLHWGSADHDGSEHIVDGVSYAAELH 120
Db 73 SNSGHSFNVDFTDENKSVLRGGLTGSLRQVHLHWGSADHDGSEHIVDGVSYAAELH 132
QY 121 VVHNSDKYPSFVEAAHEPDGLAVGLFQIGEPNSLOKIIDTLDLSIKEKGKQTRFTNF 180
```

```
Db 133 VVHNSDKYPSFVEAAHEPDGLAVGLFQIGEPNSLOKIIDTLDLSIKEKGKQTRFTNF 192
QY 181 DLLSLPPSWDYWTYPGSLTVPPLESVTWLVKQPINISSQQLAKFRSLICTAGEAAA 240
Db 193 DLLSLPPSWDYWTYPGSLTVPPLESVTWLVKQPINISSQQLAKFRSLICTAGEAAA 252
QY 241 FL 242
Db 253 FL 254
```

RESULT 5

```
US-10-099-322-304
; Sequence 304, Application US/10099322
; Publication No. US20030215449A1
; GENERAL INFORMATION:
; APPLICANT: Mezes et al.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-240CIP
; CURRENT APPLICATION NUMBER: US/10/099,322
; CURRENT FILING DATE: 2002-09-11
; PRIOR APPLICATION NUMBER: 60/261,014
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/261,018
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/318,410
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/261,013
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/261,026
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/261,029
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/313,170
; PRIOR FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: 10/044,564
; PRIOR FILING DATE: 2002-01-11
; NUMBER OF SEQ ID NOS: 324
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 304
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:consensus
; OTHER INFORMATION: sequence
US-10-099-322-304
```

```
Query Match 62.6%; Score 807; DB 15; Length 255;
Best Local Similarity 63.4%; Pred. No. 4.6e-75;
Matches 147; Conservative 30; Mismatches 55; Indels 0; Gaps 0;

QY 6 WGYREHNGPIHWKEFFFIADGQSPTEIKTEKVYDSSLRPLSIKYDPSSAKIISNSGH 65
Db 1 WGYGVHNGPEHWPPLLYIAGGDRQSPINIQTKARYDFSLKPLSVSYAATAKEINNGH 60
QY 66 SFNVDFDFTDENKSVLRGGLTGSLRQVHLHWGSADHDGSEHIVDGVSYAAELHVHWN 125
Db 61 SVQVEFDDSDMSKSVLSGGPLPAPYELKQFHFWGSSNEHSGSEHTVDGVKYPAELHLVHWN 120
QY 126 SKYPSFVEAAHEPDGLAVGLFQIGEPNSLOKIIDTLDLSIKEKGKQTRFTNFDDL 185
Db 121 STKYGSYKEAQKPDGLAVGLFVFKVGAENPGLQKLVDAIQNIKTGKSATFTNFPDSDL 180
QY 186 LPPSWDYWTYPGSLTVPPLESVTWLVKQPINISSQQLAKFRSLICTAGE 237
Db 181 LPALRDYWTYPGSLTVPPCTESVTWLVKEPITVSSEKLEKFRSLFVSVEGE 232
```

RESULT 6

```
US-10-044-564-304
; Sequence 304, Application US/10044564
```

```
; Publication No. US20040018196A1
; GENERAL INFORMATION:
; APPLICANT: Mezes et al.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-240
; CURRENT APPLICATION NUMBER: US/10/044,564
; CURRENT FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 60/261,014
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/261,018
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/318,410
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/261,013
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/261,026
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/261,029
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/313,170
; PRIOR FILING DATE: 2001-08-17
; NUMBER OF SEQ ID NOS: 306
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 304
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:consensus
; OTHER INFORMATION: sequence
US-10-044-564-304

Query Match      62.6%; Score 807; DB 15; Length 255;
Best Local Similarity 63.4%; Pred. No. 4.6e-75;
Matches 147; Conservative 30; Mismatches 55; Indels 0; Gaps 0;

QY      6  WGVREHNGPIHWKEFFPIADGQDOSPIETKTEVKYDSSLRPLSIKYDPSSAKIISNGH 65
Db      1  WGYGVHNGEHEWLLPIAGDRQSPINIQTKARVDPSLKPLSVSYAATAKEITNGH 60

QY      66  SFNVDPDDTENKSVLRGGLTGSYRLRQVHLHMGSAADHGSEHIVDGVSYAAELHVVHN 125
Db      61  SYQVEFDSDMSKSVLSGGPLPAPYRLKQPHFHWGSSNEHSGSEHTVDGVKYPAEHLVHVN 120

QY      126  SKYPSFVFAAHEPDGLAVLGVLQIGEPNSLOKIIDTLDLSIKEKGKQTRFTNFDLSL 185
Db      121  STKYSYKAEAKKPDGLAVLGVLGVKGAENPGLQKLVDAQNITKKGSAATFTNFDPSDL 180

QY      186  LPPSWDYWTYPGSLTPVPLLESVTWVLKOPINISSQOLAKFRSLLCTAEGE 237
Db      181  LPAIRDYWTYPGSLTTPCTESTVTWVLKEPITVSEQLKFRSLLFSVEGE 232

RESULT 7
US-09-802-674-4
; Sequence 4, Application US/09802674
; Patent No. US20020042088A1
; GENERAL INFORMATION:
; APPLICANT: Macina, Roberto A
; APPLICANT: Piderit, Alejandra
; APPLICANT: Sun, Yongming
; TITLE OF INVENTION: Method of Diagnosing, Monitoring, Staging, Imaging and
; TREATING Gastrointestinal Cancer
; FILE REFERENCE: DEX-0142
; CURRENT APPLICATION NUMBER: US/09/802,674
; CURRENT FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/188,061
; PRIOR FILING DATE: 2000-03-09
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 261
; TYPE: PRT

; Publication No. US20040018196A1
; GENERAL INFORMATION:
; APPLICANT: Mezes et al.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-240
; CURRENT APPLICATION NUMBER: US/10/044,564
; CURRENT FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 60/261,014
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/261,018
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/318,410
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/261,013
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/261,026
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/261,029
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/313,170
; PRIOR FILING DATE: 2001-08-17
; NUMBER OF SEQ ID NOS: 306
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 304
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:consensus
; OTHER INFORMATION: sequence
US-10-044-564-304

Query Match      62.6%; Score 807; DB 15; Length 255;
Best Local Similarity 63.4%; Pred. No. 4.6e-75;
Matches 147; Conservative 30; Mismatches 55; Indels 0; Gaps 0;

QY      6  WGVREHNGPIHWKEFFPIADGQDOSPIETKTEVKYDSSLRPLSIKYDPSSAKIISNGH 65
Db      1  WGYGVHNGEHEWLLPIAGDRQSPINIQTKARVDPSLKPLSVSYAATAKEITNGH 60

QY      66  SFNVDPDDTENKSVLRGGLTGSYRLRQVHLHMGSAADHGSEHIVDGVSYAAELHVVHN 125
Db      61  SYQVEFDSDMSKSVLSGGPLPAPYRLKQPHFHWGSSNEHSGSEHTVDGVKYPAEHLVHVN 120

QY      126  SKYPSFVFAAHEPDGLAVLGVLQIGEPNSLOKIIDTLDLSIKEKGKQTRFTNFDLSL 185
Db      121  STKYSYKAEAKKPDGLAVLGVLGVKGAENPGLQKLVDAQNITKKGSAATFTNFDPSDL 180

QY      186  LPPSWDYWTYPGSLTPVPLLESVTWVLKOPINISSQOLAKFRSLLCTAEGE 237
Db      181  LPAIRDYWTYPGSLTTPCTESTVTWVLKEPITVSEQLKFRSLLFSVEGE 232

RESULT 7
US-09-802-674-4
; Sequence 4, Application US/09802674
; Patent No. US20020042088A1
; GENERAL INFORMATION:
; APPLICANT: Macina, Roberto A
; APPLICANT: Piderit, Alejandra
; APPLICANT: Sun, Yongming
; TITLE OF INVENTION: Method of Diagnosing, Monitoring, Staging, Imaging and
; TREATING Gastrointestinal Cancer
; FILE REFERENCE: DEX-0142
; CURRENT APPLICATION NUMBER: US/09/802,674
; CURRENT FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/188,061
; PRIOR FILING DATE: 2000-03-09
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 261
; TYPE: PRT

; Publication No. US20040018196A1
; GENERAL INFORMATION:
; APPLICANT: Mezes et al.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-240
; CURRENT APPLICATION NUMBER: US/10/044,564
; CURRENT FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 60/261,014
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/261,018
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/318,410
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/261,013
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/261,026
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/261,029
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/313,170
; PRIOR FILING DATE: 2001-08-17
; NUMBER OF SEQ ID NOS: 306
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 304
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:consensus
; OTHER INFORMATION: sequence
US-10-044-564-304

Query Match      62.6%; Score 807; DB 15; Length 255;
Best Local Similarity 63.4%; Pred. No. 4.6e-75;
Matches 147; Conservative 30; Mismatches 55; Indels 0; Gaps 0;

QY      6  WGVREHNGPIHWKEFFPIADGQDOSPIETKTEVKYDSSLRPLSIKYDPSSAKIISNGH 65
Db      1  WGYGVHNGEHEWLLPIAGDRQSPINIQTKARVDPSLKPLSVSYAATAKEITNGH 60

QY      66  SFNVDPDDTENKSVLRGGLTGSYRLRQVHLHMGSAADHGSEHIVDGVSYAAELHVVHN 125
Db      61  SYQVEFDSDMSKSVLSGGPLPAPYRLKQPHFHWGSSNEHSGSEHTVDGVKYPAEHLVHVN 120

QY      126  SKYPSFVFAAHEPDGLAVLGVLQIGEPNSLOKIIDTLDLSIKEKGKQTRFTNFDLSL 185
Db      121  STKYSYKAEAKKPDGLAVLGVLGVKGAENPGLQKLVDAQNITKKGSAATFTNFDPSDL 180

QY      186  LPPSWDYWTYPGSLTPVPLLESVTWVLKOPINISSQOLAKFRSLLCTAEGE 237
Db      181  LPAIRDYWTYPGSLTTPCTESTVTWVLKEPITVSEQLKFRSLLFSVEGE 232

RESULT 7
US-09-802-674-4
; Sequence 4, Application US/09802674
; Patent No. US20020042088A1
; GENERAL INFORMATION:
; APPLICANT: Macina, Roberto A
; APPLICANT: Piderit, Alejandra
; APPLICANT: Sun, Yongming
; TITLE OF INVENTION: Method of Diagnosing, Monitoring, Staging, Imaging and
; TREATING Gastrointestinal Cancer
; FILE REFERENCE: DEX-0142
; CURRENT APPLICATION NUMBER: US/09/802,674
; CURRENT FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/188,061
; PRIOR FILING DATE: 2000-03-09
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 261
; TYPE: PRT
```

```
; ORGANISM: Homo sapiens
US-09-802-674-4

Query Match      62.4%; Score 804; DB 9; Length 261;
Best Local Similarity 59.4%; Pred. No. 9.7e-75;
Matches 142; Conservative 44; Mismatches 53; Indels 0; Gaps 0;

QY      1  MSRLSWGVEHNGPIHWKEFFPIADGQDOSPIETKTEVKYDSSLRPLSIKYDPSSAKII 60
Db      1  MASPDWGYDDKNGPEQMSKLYPIANGNNSPVDIKTSETKHDTLSLAPISVSNPATAKEI 60

QY      61  SNSGHSFNVDFTDENKSVLRGGLTGSYRLRQVHLHMGSAADHGSEHIVDGVSYAAELH 120
Db      61  INVGHSHVNFEDNDNRSVLKGPFSDSYRLQFHFHMGSTNEHSGSEHTVDGVKYSAEHL 120

QY      121  VVHNSDKYPSFVFAAHEPDGLAVLGVLQIGEPNSLOKIIDTLDLSIKEKGKQTRFTNF 180
Db      121  VAHNSAKYSSSLAAEAKADGLAVLGVLKMGVEANPKLQKLVDAQIAIKTKGRAPFTNF 180

QY      181  DLLSLLPPSWDYWTYPGSLTPVPLLESVTWVLKOPINISSQOLAKFRSLLCTAEGEAA 239
Db      181  DFTLLPSSLDFTWYTPGSLTHPPLYESVTWIIICKESISVSSQLAQFRSLLSNVEGDNA 239

RESULT 8
US-09-981-353-80
; Sequence 80, Application US/09981353
; Patent No. US20020160382A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy W.
; APPLICANT: Jones, David A.
; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
; FILE REFERENCE: PA-0038 US
; CURRENT APPLICATION NUMBER: US/09/981,353
; CURRENT FILING DATE: 2001-10-11
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 80
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020160382A1 2101663CD1
US-09-981-353-80

Query Match      62.4%; Score 804; DB 9; Length 261;
Best Local Similarity 59.4%; Pred. No. 9.7e-75;
Matches 142; Conservative 44; Mismatches 53; Indels 0; Gaps 0;

QY      1  MSRLSWGVEHNGPIHWKEFFPIADGQDOSPIETKTEVKYDSSLRPLSIKYDPSSAKII 60
Db      1  MASPDWGYDDKNGPEQMSKLYPIANGNNSPVDIKTSETKHDTLSLAPISVSNPATAKEI 60

QY      61  SNSGHSFNVDFTDENKSVLRGGLTGSYRLRQVHLHMGSAADHGSEHIVDGVSYAAELH 120
Db      61  INVGHSHVNFEDNDNRSVLKGPFSDSYRLQFHFHMGSTNEHSGSEHTVDGVKYSAEHL 120

QY      121  VVHNSDKYPSFVFAAHEPDGLAVLGVLQIGEPNSLOKIIDTLDLSIKEKGKQTRFTNF 180
Db      121  VAHNSAKYSSSLAAEAKADGLAVLGVLKMGVEANPKLQKLVDAQIAIKTKGRAPFTNF 180

QY      181  DLLSLLPPSWDYWTYPGSLTPVPLLESVTWVLKOPINISSQOLAKFRSLLCTAEGEAA 239
Db      181  DFTLLPSSLDFTWYTPGSLTHPPLYESVTWIIICKESISVSSQLAQFRSLLSNVEGDNA 239

RESULT 9
US-10-235-994-24
; Sequence 24, Application US/10235994
; Publication No. US20030101002A1
; GENERAL INFORMATION:
; APPLICANT: Bartha, Gabor
```



```
;
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: Gene
; LOCATION: (1)..(260)
; OTHER INFORMATION: Carbonic anhydrase domain of human carbonic anhydrase I
US-09-983-000A-26

Query Match          62.2%; Score 802; DB 10; Length 260;
Best Local Similarity 60.3%; Pred. No. 1.6e-74;
Matches 141; Conservative 43; Mismatches 50; Indels 0; Gaps 0;

QY 6 WGYREHNGPIHWKEFFPIADGQOQSPIEIKTEKVYDSSLPLSIKYDPSSAKIISNGH 65
Db 5 WGYDDKNGPEQWKLPIANGNNQSPVDIKTSEYKHDTSKPISVSNPATAKEIINVGH 64

QY 66 SFNVDFDTEKNSVLRGGPLTGSYRLRVHLHWSADHGHSEHIVDGVSYAAELHVVHWN 125
Db 65 SFHNVFENDNKRSLVLRGGPFSDSYRLRFQHFHWGSTNEHSEHTVDGVKYSALHVAHWN 124

QY 126 SDKYPFVEAAHEPDGLAVLGVLQIGEPNSQLQKITDTLDSIKEKGKQTRFTNFDLLSL 185
Db 125 SAKYSSLAFAASKADGLAVIGVLMKVGEANPKLQKVLDAQAIAKTKGRAPTFNFDPSL 184

QY 186 LPPSDYWTYPGSLTPVPLESVTWIVLKQPINISSQOLAKFRSLICTAEGEA 239
Db 185 LPSSLDFTYTPGSLTHPLYESVTWICKESISVSSEQLAQFRSLLSNVEGDNA 238
```

```
RESULT 13
US-10-000-954-4
; Sequence 4, Application US/10000954
; Publication No. US20020127226A1
; GENERAL INFORMATION:
; APPLICANT: Schlesinger, Joseph
; Barnea, Gilad
; Grumet, Martin H.
; Margolis, Richard U.
; TITLE OF INVENTION: A NEW CLASS OF RPTASES: THEIR
; STRUCTURAL DOMAINS AND LIGANDS
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/10/000,954
; FILING DATE: 04-Dec-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/644,293
; FILING DATE: 23-Aug-2000
; APPLICATION NUMBER: 08/081,929
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30742
; REFERENCE/DOCKET NUMBER: 7683-041-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 869-8864/9741
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 260 amino acids
```

```
;
; LENGTH: 260 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-10-000-954-4

Query Match          62.2%; Score 802; DB 13; Length 260;
Best Local Similarity 60.3%; Pred. No. 1.6e-74;
Matches 141; Conservative 43; Mismatches 50; Indels 0; Gaps 0;

QY 6 WGYREHNGPIHWKEFFPIADGQOQSPIEIKTEKVYDSSLPLSIKYDPSSAKIISNGH 65
Db 5 WGYDDKNGPEQWKLPIANGNNQSPVDIKTSEYKHDTSKPISVSNPATAKEIINVGH 64

QY 66 SFNVDFDTEKNSVLRGGPLTGSYRLRVHLHWSADHGHSEHIVDGVSYAAELHVVHWN 125
Db 65 SFHNVFENDNKRSLVLRGGPFSDSYRLRFQHFHWGSTNEHSEHTVDGVKYSALHVAHWN 124

QY 126 SDKYPFVEAAHEPDGLAVLGVLQIGEPNSQLQKITDTLDSIKEKGKQTRFTNFDLLSL 185
Db 125 SAKYSSLAFAASKADGLAVIGVLMKVGEANPKLQKVLDAQAIAKTKGRAPTFNFDPSL 184

QY 186 LPPSDYWTYPGSLTPVPLESVTWIVLKQPINISSQOLAKFRSLICTAEGEA 239
Db 185 LPSSLDFTYTPGSLTHPLYESVTWICKESISVSSEQLAQFRSLLSNVEGDNA 238
```

```
RESULT 14
US-10-000-954-5
; Sequence 5, Application US/10000954
; Publication No. US20020127226A1
; GENERAL INFORMATION:
; APPLICANT: Schlesinger, Joseph
; Barnea, Gilad
; Grumet, Martin H.
; Margolis, Richard U.
; TITLE OF INVENTION: A NEW CLASS OF RPTASES: THEIR
; STRUCTURAL DOMAINS AND LIGANDS
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/10/000,954
; FILING DATE: 04-Dec-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/644,293
; FILING DATE: 23-Aug-2000
; APPLICATION NUMBER: 08/081,929
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30742
; REFERENCE/DOCKET NUMBER: 7683-041-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 869-8864/9741
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 259 amino acids
```

```
;
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-10-000-954-5

Query Match      61.2%; Score 788.5; DB 13; Length 259;
Best Local Similarity 61.6%; Pred. No. 3.9e-73;
Matches 143; Conservative 33; Mismatches 55; Indels 1; Gaps 1;

QY 6 WGYREHNGPIHWKEFPPIADGQQQPIETKTEVKYDSSLRPLSIKYDPSSAKIISNSGH 65
Db 4 WGYGKHNGPEHWHKDFPIAKGERQSPVDITHTAKYDPSLKPLSVSYDOATSLRIINNGH 63

QY 66 SFNVDFDPTENKSVLRGGPLTGYRLRQVHLHWGSADHDGSEHIYDGVSYAAELHVVHWN 125
Db 64 AFNVEFDDSDQKAVLKGGLDGTYRLIQPHFWHWSLDGGSEHTVDKKKYAAELHLVHWN 123

QY 126 SDKYPSFVEAAHEPDLAVLGVFLQIGEPNSQLKITDITLDSIKKKGKQTRFTNFDLLSL 185
Db 124 T-KYDGFKAQQPQDGLAVLGIFLKVGSAPKGLQKVVDVLDISIKTKGSADFTNFDPRGL 182

QY 186 LPPSWDYWTYPGSLTVPPLLESVTWIVLKQPINISSQQLAKFRSLICTAEGE 237
Db 183 LPESLDYWTYPGSLTTPPLLECCTWIVLKEPISVSSEQVLKPKRLNFNGEGE 234

RESULT 15
US-09-981-353-117
; Sequence 117, Application US/09981353
; Patent No. US20020160382A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy W.
; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
; FILE REFERENCE: PA-0038 US
; CURRENT APPLICATION NUMBER: US/09/981,353
; CURRENT FILING DATE: 2001-10-11
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 117
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020160382A1 2742913CD1
US-09-981-353-117

Query Match      61.2%; Score 788.5; DB 9; Length 260;
Best Local Similarity 61.6%; Pred. No. 3.9e-73;
Matches 143; Conservative 33; Mismatches 55; Indels 1; Gaps 1;

QY 6 WGYREHNGPIHWKEFPPIADGQQQPIETKTEVKYDSSLRPLSIKYDPSSAKIISNSGH 65
Db 5 WGYGKHNGPEHWHKDFPIAKGERQSPVDITHTAKYDPSLKPLSVSYDOATSLRIINNGH 64

QY 66 SFNVDFDPTENKSVLRGGPLTGYRLRQVHLHWGSADHDGSEHIYDGVSYAAELHVVHWN 125
Db 65 AFNVEFDDSDQKAVLKGGLDGTYRLIQPHFWHWSLDGGSEHTVDKKKYAAELHLVHWN 124

QY 126 SDKYPSFVEAAHEPDLAVLGVFLQIGEPNSQLKITDITLDSIKKKGKQTRFTNFDLLSL 185
Db 125 T-KYDGFKAQQPQDGLAVLGIFLKVGSAPKGLQKVVDVLDISIKTKGSADFTNFDPRGL 183

QY 186 LPPSWDYWTYPGSLTVPPLLESVTWIVLKQPINISSQQLAKFRSLICTAEGE 237
Db 184 LPESLDYWTYPGSLTTPPLLECCTWIVLKEPISVSSEQVLKPKRLNFNGEGE 235
```

Blank
Sheet